NTNU – Trondheim
Norwegian University of
Science and Technology

Structural and biochemical insights into mammalian cobalt-substituted methionine sulfoxide reductase B1 using UV-visible spectroscopy and high-resolution NMR spectroscopy

Henrik Waldal Holen

Biotechnology

Supervisor: Oleksandr Dykyy, IBT

Norwegian University of Science and Technology
Department of Biotechnology

Abstract

Prior to this study it was discovered that MsrB1 from *Mus musculus* expressed in *Escherichia Coli* binds cobalt(II) (hereafter cobalt) in cobalt-supplemented growth media, and it had further been demonstrated that the His-tag was not responsible for this metal uptake. The aim of this study was to investigate the effects of cobalt on the growth of *E. Coli* in culture, characterize the metal-uptake and metal-binding site of cobalt-substituted MsrB1 by UV-visible spectroscopy and to gain structural information about the protein by high resolution NMR spectroscopy.

The effects of cobalt on growth of *E. Coli* were studied by growing cultures in Lysogeny broth (LB) and minimal (M9)-media supplemented with different concentrations of cobaltdichloride ($CoCl_2$) and monitoring culture growth by optical density (OD) measurements. Growth rates were found to decrease with increasing concentrations of $CoCl_2$.

To study the cobalt-uptake of MsrB1, the protein was recombinantly expressed in *E. Coli* in different cobalt-supplemented growth media, and the purified protein analyzed by UV-vis spectroscopy. Cobalt-uptake was demonstrated in all cases by characteristic absorption peaks owing to the cobalt-ligand complex, and the wavelengths of these peaks matched those of a tetrahedral four-coordinated cobalt-Cys complex. It was argued that these four Cys residues should be the same that constitute the zinc(II)-binding site of MsrB1, indicating that cobalt simply replaces zinc as a structural metal ion in cobalt-substituted MsrB1 (Co-MsrB1). It was argued that the intracellular concentration of cobalt in *E. Coli* should be significantly higher than zinc, and that this together with similar ionic radii for cobalt and zinc leads to the formation of Co-MsrB1. MsrB1 expressed in nickel(II)-supplemented LB did not lead to formation of Ni-MsrB1, which was argued to result from nickel not being released directly into the cytosol in *E. Coli*.

Co-MsrB1 was produced by zinc-starvation of *E. Coli* followed by expressing the protein in zinc-free minimal medium supplemented with $CoCl_2$. To investigate if pH-titratable groups could be detected, the protein was dialyzed against buffers with pH 4.9-11.5 and the molar extinction coefficient was found from UV-vis absorption spectra in the different pH. Two titration curves were observed, but assignment of the titrations to specific residues could not be made. Further, Co- and Zn-MsrB1 was dialyzed against buffers with metal chelating agents to remove the metal ions from the two proteins. Cobalt was successfully removed at pH 5.0 and 5.5, while removal of zinc from Zn-MsrB1 was not detected, demonstrating that zinc is more tightly bound to the Cys-ligands than cobalt.

To study the ratio of formation of Co-MsrB1 and the native zinc-form Zn-MsrB1 in $CoCl_2$-supplemented growth media, the molar extinction coefficient of Co-MsrB1 was determined,

and the concentration of Co-MsrB1 in purified protein samples from protein expression in different growth media was determined. The Co-MsrB1:Zn-MsrB1 ratio was found to be 0.2 in M9 medium supplemented with 10 µM CoCl$_2$, 0.1 in LB supplemented with 50 µM CoCl$_2$ and 0.03 in LB supplemented with 10 µM CoCl$_2$.

From 2D- and 3D-NMR experiments on ^{13}C- and ^{15}N-enriched Co-MsrB1, a 70 % backbone assignment and 50 % side chain-assignment was accomplished using Computer Aided Resonance Assignment (CARA). The four structural Cys-residues of the native protein was not found, while the other three Cys-residues of MsrB1 were assigned, confirming that the same Cys-residues are responsible for coordination of cobalt and zinc in MsrB1. Many strongly shifted signals were observed in 1D ^{1}H spectra of Co-MsrB1, some as far upfield as 350 ppm and downfield as -80 ppm, and it was argued that most of the unassigned residues should be found outside the spectral width of the 2D and 3D-NMR spectra.

To gain further structural information about Co-MsrB1, pseudocontact shifts (PCSs) were determined for the assigned H^N-atoms of Co-MsrB1 by using the published chemical shifts of for Zn-MsrB1. The PCSs were analyzed by AnisoFit using three conformers and the mean conformer of the published Zn-MsrB1 structure, and the best correlations between observed and calculated PCSs were found for conformer 3. The PCSs were plotted against their hypothetical distance to cobalt using the structure of Zn-MsrB1, and a very good PCS-distance proportionality was found, indicating that Co-MsrB1 and Zn-MsrB1 have the same overall fold and structure. The AnisoFit calculations and PCSs of the N-terminus suggested that the N-terminus spends significant time in the proximity of the metal-binding site, and it was argued that this proximity ensures high catalytic efficiency due to the short distance between the catalytic and resolving Cys-residues.

Sammendrag

På forhånd av denne studien ble det oppdaget at proteinet MsrB1 fra *Mus musculus* uttrykket i *Escherichia Coli* binder kobolt(II) (heretter kobolt) i kobolt-supplert vekstmedium, og det hadde videre blitt demonstrert at det ikke er His-tag-sekvensen som er ansvarlig for å binde metallionet. Målet med denne studien var å undersøke effektene av kobolt på vekst i *E. Coli*-kulturer, karakterisere metall-opptaket og metall-bindesetet i kobolt-substituert MsrB1 ved UV-visible absorpsjonsspekroskopi og erverve strukturell informasjon om proteinet ved høyoppløsning NMR-spektroskopi.

Effektene av kobolt på vekst av *E. Coli* ble studert ved å dyrke kulturer i Lysogeny broth (LB) og minimalt (M9)-medium supplert med forskjellige konsentrasjoner av koboltdiklorid (CoCl$_2$) og overvåking av vekst ved målinger av optisk tetthet. Det ble vist at vekstratene sank med økende konsentrasjoner av CoCl$_2$ i mediet.

For å studere kobolt-opptaket til MsrB1 ble proteinet rekombinant uttrykt i *E. coli* i ulike kobolt-supplerte vekstmedier, og det purifiserte proteinet ble analysert ved hjelp av UV-vis spektroskopi. Kobolt-opptaket ble demonstrert i alle tilfeller av karakteristiske absorpsjonstopper som tilhørte kobolt-ligand-komplekset, og bølgelengdene til disse toppene passet til et tetrahedralt fire-koordinert kobolt-Cys-kompleks. Det ble argumentert for at disse Cys-aminosyrene er de samme som utgjør sink(II)-bindesetet i MsrB1, som indikerer at kobolt simpelthen erstatter sink som strukturelt metallion i kobolt-substitutert MsrB1 (Co-MsrB1). Det ble argumentert for at den intracellulære konsentrasjonen av kobolt i *E. Coli* er mye høyere enn for sink, og at dette sammen med like ioneradiuser for kobolt og sink fører til dannelse av Co-MsrB1. MsrB1 uttrykt i nikkel(II)-supplert LB ledet ikke til dannelse av Ni-MsrB1, og det ble argumentert for at dette skyldes at nikkel ikke slippes direkte inn i cytosol hos *E. Coli*.

Co-MsrB1 ble produsert ved sink-sulting av *E. Coli* etterfulgt av uttrykk av proteinet i sinkfritt M9-vekstmedium supplert med CoCl$_2$. For å undersøke om pH-titrerbare grupper kunne detekteres ble proteinet dialysert mot buffere med pH 4.9-11.5 og den molare ekstinksjonskoeffisienten ble funnet fra UV-vis absorpsjonsspektra i ulik pH. To titreringskurver ble observert, men kunne ikke tilskrives noen spesifikke aminosyrer. Videre ble Co- og Zn-MsrB1 dialysert mot buffere med metallkelaterende forbindelser for å fjerne metallionene fra de to proteinene. Kobolt ble påvist fjernet ved pH 5.0 og 5.5, mens sink ikke lot seg fjerne, noe som viste at sink er sterkere bundet til Cys-ligandene enn kobolt.

For å studere dannelsesforholdet mellom Co-MsrB1 og det naturlige Zn-MsrB1 i CoCl$_2$-supplert vekstmedia ble den molare ekstinksjonskoeffisienten for Co-MsrB1 fastslått, og fra dette ble konsentrasjonen av Co-MsrB1 i de purifiserte proteinprøvene fra uttrykk i de ulike media fastslått. Co-MsrB1:Zn-MsrB1 forholdet ble funnet å være omtrent 0.2 i M9-medium

supplert med 10 µM CoCl$_2$, 0.1 i LB supplert med 50 µM CoCl$_2$ og 0.03 i LB supplert med 10 µM CoCl$_2$.

Fra 2D- og 3D-NMR-eksperimenter på [13]C- og ^{15}N-merket Co-MsrB1 ble det oppnådd et 70 % ryggrad-assignment og 50 % sidekjede-assignment ved hjelp av Computer Aided Resonance Assignment (CARA). De fire structurelle Cys-aminosyrene ble ikke funnet, mens de andre tre Cys-aminosyrene til MsrB1 ble assignet, hvilket bekreftet at de samme Cys koordinerer kobolt og sink i MsrB1. Mange sterkt skiftede signaler ble observert i 1D [1]H-spektra av Co-MsrB1, noen så langt opp som 350 ppm og ned som -80 ppm, og det ble argumentert for at de fleste av de uassignede aminosyrene er skiftet ut av spektralvinduene til 2D- og 3D-NMR-spektrene.

For å få videre strukturell informasjon om Co-MsrB1 ble pseudokontakt-skift (PCS) funnet for de assignede H[N]-atomene i Co-MsrB1 ved å bruke de publiserte kjemiske skiftene for Zn-MsrB1. PCSene ble analysert ved hjelp av AnisoFit ved å bruke tre konformer og gjennomsnittskonformeren til den publiserte Zn-MsrB1-strukturen, og de beste korrelasjonene mellom observert og kalkulert PCSs ble funnet for konformer 3. PCSene ble plottet mot deres hypotetiske avstand til kobolt ved hjelp av Zn-MsrB1-strukturen som modell, og en veldig god avstand-PCS-proporsjonalitet ble funnet, hvilket viste at Co-MsrB1 og Zn-MsrB1 har samme folding og struktur. AnisoFit-kalkulasjonene og PCSene til N-terminusen antydet at N-terminusen tilbringer signifikant tid i nærheten av kobolt-bindesetet, og det ble argumentert for at denne nærheten sikrer høy katalytisk aktivitet på grunn av lav avstand mellom den katalytiske og løsende Cys-aminosyren.

Preface

This thesis was written at the Department of Biotechnology of the Norwegian University of Science and Technology (NTNU) and is the result of a 2-year master program in biotechnology (Master of Science).

Table of Contents

1 Theory

1.1 Protein structure – building blocks and the basis for structure

This section is adapted from [1] and [2]. All figures are reproduced from [2].

The unique biological and biochemical properties of individual proteins as well as their secondary, tertiary and quaternary structures are largely determined by their amino acid sequence (also called primary structure). Below, the basic principles by which protein structure are determined is described.

1.1.1 Amino acids are the primary building blocks

Proteins are synthesized in a cell using messenger RNA - a transcript of DNA - as a template, thereby translating the genetic sequence into a gene product. In this process, amino acids are joined end-to-end by peptide bonds to form a specific sequence determined by the RNA transcript. This is referred to as the primary structure.

Amino acids have a central carbon atom (C_α) connected to a hydrogen atom (H), an amino group (NH_3^+), a carboxyl group (COOH) and a side-chain (R) (Figure. 1.1).

$$H_3\overset{+}{N}-\underset{R}{\overset{COO^-}{\underset{|}{\overset{|}{C}}}}-H$$
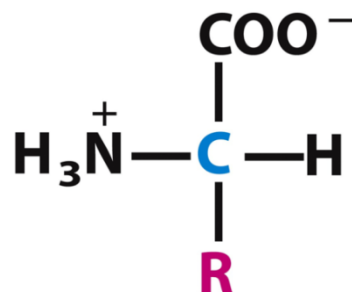
Figure 1.1: General structure of amino acids. Schematic representation of an amino acid, where C_α is colored blue while the side chain is denoted R and colored red. For proline, the side chain is connected to C_α by two bonds, and there is therefore no H-atom bound to C_α.

The C_α is a chiral center, and an amino acid therefore has two enantiomers, the L- and D-form, the former of which is the prevalent form in nature (with some exceptions). Different amino acids have different side chains, and there are 20 common amino acids. These are divided into polar, non-polar/hydrophobic and charged amino acids. Glycine (Gly) and proline (Pro) represent special cases, the former having a hydrogen atom instead of a side-chain and the latter having a side-chain connected to its amino group nitrogen atom. The

specific sequence of amino acids is referred to as the primary structure of a protein or polypeptide.

1.1.2 Phi and psi bond angles are the basis of structure

The peptide bonds have a partial double-bond character, and thus rotation around this bond will not happen under normal conditions. The C_α-C- and N-C_α-bonds, however, are single bonds which allow rotation. One therefore generally divides a polypeptide chain into peptide units (or groups), each unit consisting of the peptide-bond atoms (carbonyl- and amide group) and the two flanking C_α-atoms, and the atoms of the peptide units are also referred to as the backbone of the polypeptide. Each C_α-atom belongs to two peptide groups. Because of the rotational freedom around the C_α-atoms of the backbone, each peptide group is characterized by two bond angles, namely the φ- (phi) and ψ- (psi) dihedral angles. The φ-dihedral angle is defined by the plane of the peptide bond and rotation around the N-C_α-bond, and the ψ-dihedral angle is defined by the plane of the peptide bond and the rotation around the C_α-C-bond (Figure. 1.2).
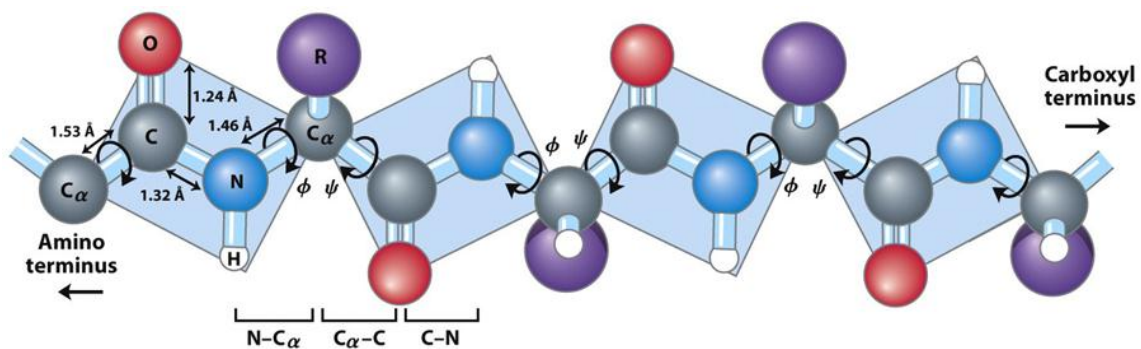
Figure 1.2: Rotational bonds in a polypeptide chain. Schematic representation of a polypeptide, with the planes of the peptide groups illustrated in blue. The bonds which allow rotation are indicated with circular arrows, and denoted φ (phi) or ψ (psi) to indicate which angles the rotation affects.

Not all dihedral angle combinations for are energetically favored. This is due to steric hindrance between the atoms of the peptide units (the backbone atoms) and the side chains of the amino acids (Figure 1.3). Because the different amino acids have different side chains, the dihedral angle ranges that are allowed also differ between the amino acids. Two extreme cases are Gly and Pro. Gly, having a H-atom as side chain, can assume a relatively broad range of conformations. Pro, on the other hand, with a side chain connected to the N-atom of the backbone, is highly restricted.
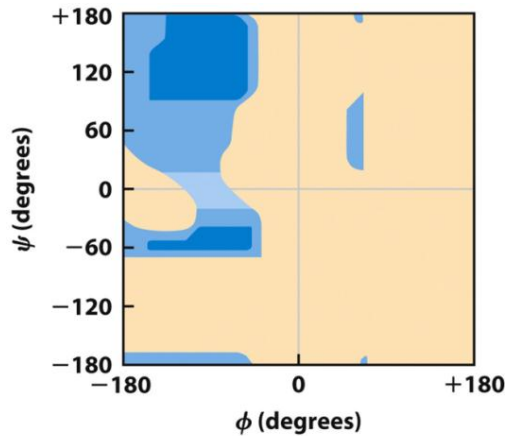
Figure 1.3: Ramachandran plot showing allowed dihedral angles. Dihedral angles shown as a two-dimensional diagram, also known as a Ramachandran plot. Regions of favored dihedral angles are shown in blue. Darker blue regions indicate the dihedral angles with the least steric hindrance. Yellow/beige regions indicate dihedral angles that are less favored or not allowed. Conformations outside of the indicated allowed regions occur. For Gly, the area of allowed dihedral angles is bigger than indicated in this diagram. Pro is much more restricted.

1.1.3 Secondary structure

Relatively fixed dihedral angles of the peptide units give rise to structural elements within a protein. For a given stretch of amino acids, this is called secondary structure. Since only certain conformations are allowed for most amino acids, the number of stable secondary structure types are limited. The two most common secondary structures are the α-helix and the β-strand. Both structures are associated with certain dihedral angles (Figure 1.4).
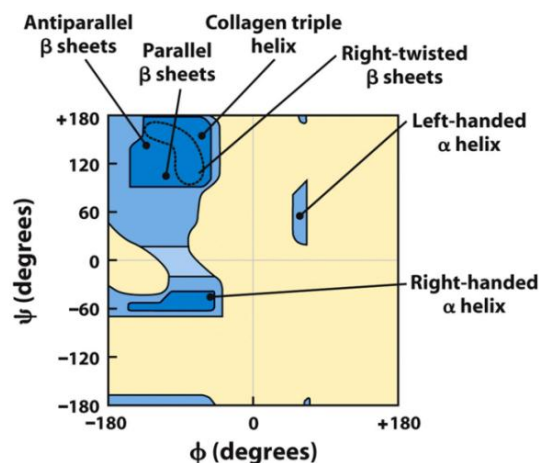
Fig. 1.4: Dihedral angles and secondary structure. Ramachandran plot showing how different dihedral angles are associated with different secondary structures. α-helices and β-sheets are explained further below in the discussion. Left-handed α-helices are theoretically possible but have not yet been observed to exist.

3

In the α-helix, the backbone coils around an imaginary axis and thus forms a helical structure, typically with about 3.6 amino acids per turn of the coil. The conformations are stabilized by hydrogen bonds between the carbonyl oxygen and amino-hydrogen of the backbone, and the amino acid side chains protrude from the helix (Figure 1.5).
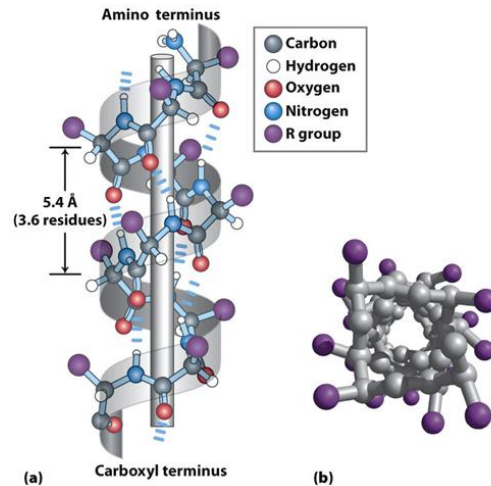
Figure 1.5: The α-helix. a) schematic representation of an α-helix with a grey ribbon illustrating the helical turns and the imaginary axis shown as a gray rod. Hydrogen bonds are shown stipulated in blue. b) Ball-and-stick model of an α-helix as seen from above through its center.

In an α-helix, all the amino acids exhibit similar dihedral angles, resulting in the repeating helical turns. Amino acids that cannot assume these conformations or that have large, bulky side chains will destabilize the helix or introduce a kink. In particular, Pro and Gly are the amino acids least likely to occur in a stable α-helix. Pro is unable to rotate around the N-C$_\alpha$-bond due to the side chain being connected to the nitrogen-atom of the backbone while Gly, having very high flexibility, normally assumes other dihedral angles than what is typical in an α-helix. The α-helix is predominantly right-handed, owing to the low stability of left-handed helices with L-amino acids.

β-strands, or β-conformations, form when a stretch of amino acids assume the most extended conformations. This results in a zigzag-like structure of the backbone, with the side-chains protruding to opposite sides for every second residue. β-strands can be packed together side by side to form β-sheets. Like in the α-helix, hydrogen bonds between the atoms of the backbone stabilize β-sheets. When β-strands that run in the same direction are packed together in a β-sheet (defined by the amino- and carboxy terminus of the protein), it is called a parallel β-sheet. When formed by strands running in the opposite direction it is called an antiparallel β-sheet (Figure. 1.6). β-sheets can also consist of both parallel and antiparallel packing and is then referred to as a mixed β-sheet.
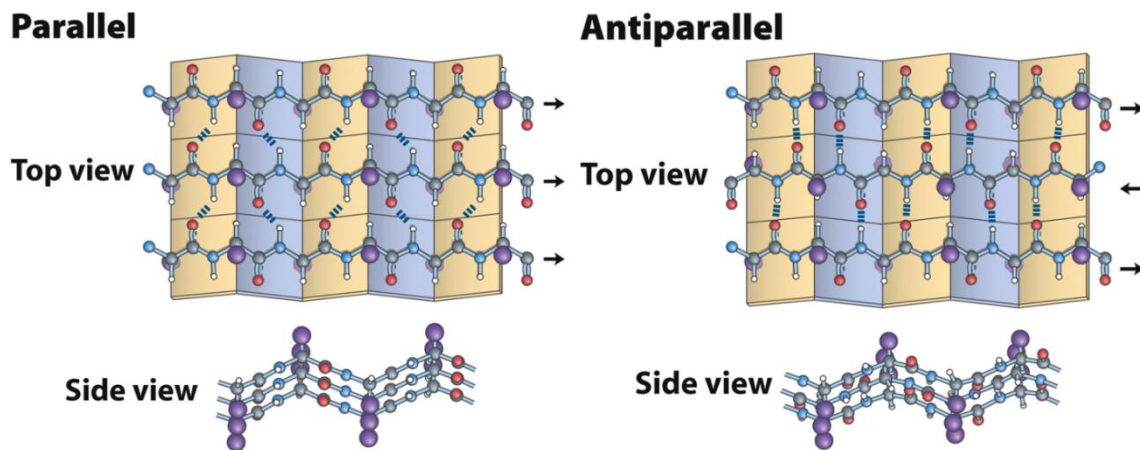
Figure 1.6: The β-sheet. β-strands packed together to form β-sheets, with parallel or antiparallel alignment. The direction of the polypeptide chain is shown with black arrows, hydrogen bonds are stipulated in blue and yellow/blue background shows the zigzag-like plane of the β-sheet (top view). Packing in parallel and antiparallel β-sheets are different due to the different directions of the strands (top view and side view).

1.1.4 Tertiary structure

The overall three-dimensional arrangement of the atoms in a protein is called its tertiary structure. If a protein consists of two or more individual proteins (subunits) in a complex, one refers to the overall structure as the quaternary structure (Figure 1.7). Typically, a protein consists of several secondary structure elements (α-helices and β-strands) connected by stretches of amino acids called loops or loop regions, which vary in length. Loop regions may be very flexible, thus lacking a defined structure, or more rigid, depending on both its length and function within the protein. Certain simple combinations of secondary structure elements can form motifs or domains that are common for many proteins. of β-sheets is an example such a structural motif. The tertiary structure of a protein may contain of one or more such motifs, and the combination and spatial arrangement of motifs and/or secondary structure elements is referred to as the fold of a protein. The fold may belong to a class of folds shared by different (often evolutionary related) proteins, or it may be unique to the specific protein.

Proteins that belong to the cytosol typically have residues with hydrophobic side chains tightly packed towards the inside, thus creating a hydrophobic core, while the hydrophilic residues (i.e. polar or charged side chains) face towards the outside. This is the main driving force for polypeptide chain in the cytosol to fold into a stable, structured protein. For some proteins that are embedded in the cell membrane, the opposite is often true. Here, α-helices may be associated and form a hydrophilic core and have hydrophobic residues facing the outside (the lipid phase).
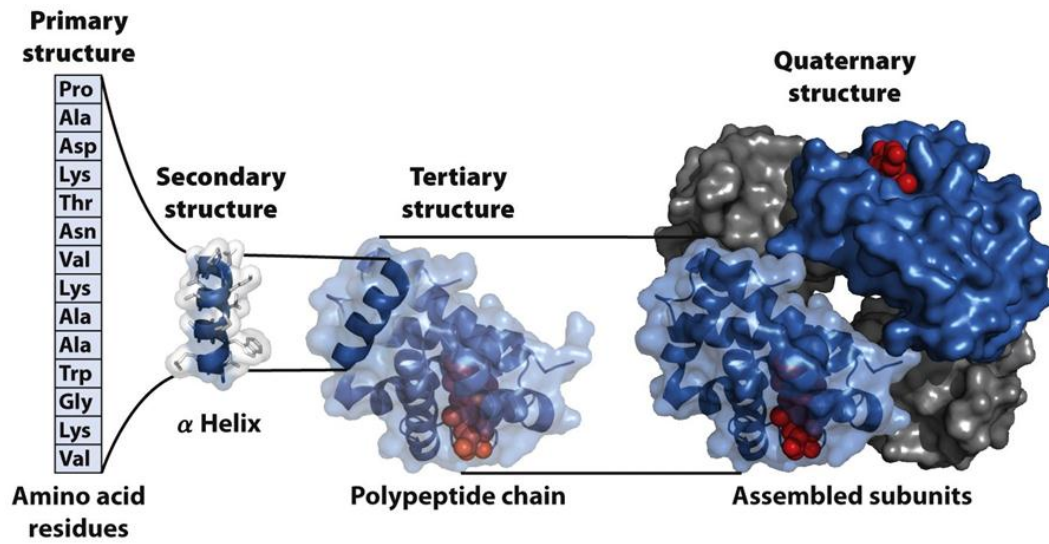
Figure 1.7: Different elements of protein structure. Illustration of the protein structure of hemoglobin at different the levels, from primary to quaternary structure.

1.2 Metal binding sites in proteins

1.2.1 Metals in biology – distribution and function

Many metal ions are essential to living cells, and most of them are so-called trace elements (elements that are needed by an organism in minute amounts) [2] (Figure 1.8).

Fig. 1.8: Elements essential to life: Essential elements with their respective places in the periodic table. Essential bulk elements, of which organisms need relatively large amounts, are highlighted in beige. Essential trace elements are highlighted in yellow. As can be seen from the table, most trace elements are transition metals. Figure modified from [2].

In addition to other biological functions, metal ions are crucial to the structure and/or function of many proteins [3]. Occurrence of different metal ions as enzymatic cofactors is shown in Figure 1.9.
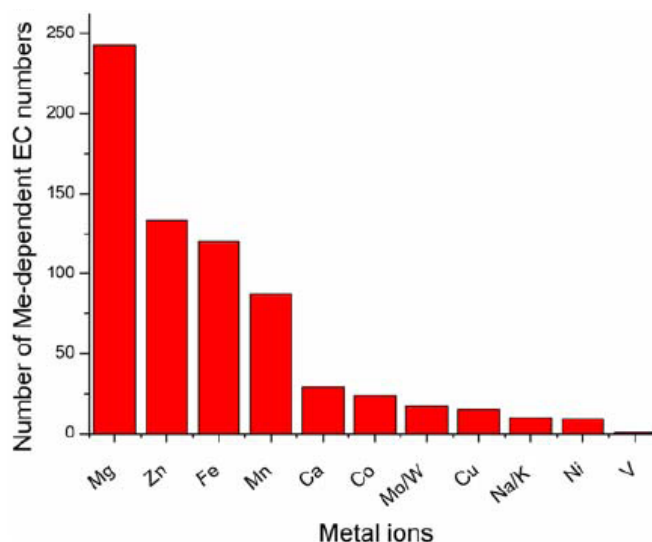
Figure 1.9: Distribution of metals as cofactors. Occurrence of various metal ions as cofactors in enzymes with known structure, the bars indicating the number of distinct enzymes (EC number) which depend on the given metal ion for catalysis. Figure reproduced from [3].
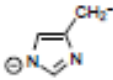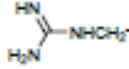
Proteins that require metal ions have a metal binding site. The roles of metal binding sites can be divided into five categories [4]:

1. Structural – bound metal forms a structural motif and/or assists correct folding
2. Metal storage – uptake, binding and release of metals in soluble form
3. Electron transfer – bound metal facilitates uptake, release and storage of electrons
4. Dioxygen binding – bound metal facilitates binding, storage and release of molecular oxygen
5. Catalytic – bound metal facilitates substrate binding, activation and turnover

1.2.2 Metal ligands in proteins

In a metal binding site, metal ions are either bound to endogenous ligands (provided by backbone or side chain atoms of the same polypeptide) or exogenous ligands (provided by other molecules bound to the polypeptide) [4]. A list of endogenous ligands is shown in Table 1.1.

Table 1.1: List of endogenous ligands. Endogenous ligands divided into three groups by which atom is donating electrons to the metal ion. In the left column the different ligand groups are listed, with examples of these groups listed in the right column. X represents. any residue unless specific residues are listed in brackets. Table modified from [4].

As can be seen from Table 1.1, most endogenous ligands are proton donors/acceptors. Metal ions are normally bound to endogenous ligands by covalent bonds, and the strength

of the metal-ligand bond is determined by a metal ions ability to compete with the proton in binding. A strong metal-ligand bond thus lowers the pK$_a$ of the ligand significantly [4].

Many protein-bound metals are divalent ions. Affinity of ligands for divalent first-row transition metals generally follows the trend

$$Ca^{2+} < Mg^{2+} < Mn^{2+} < Fe^{2+} < Co^{2+} < Zn^{2+} < Ni^{2+} < Cu^{2+}$$
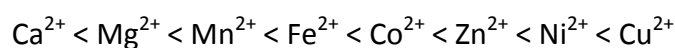
also known as the Irving-Williams series. The trend follows a decrease in ionic radii, which leads to stronger metal-ligand bonds. Zn^{2+} is larger than Cu^{2+}, and is thus an exception. It is assumed that metal binding sites of most polypeptides formed *in vivo* bind acquire metals by spontaneous self-assembly. This is supported by the fact that certain inorganic metal-centers (e.g. Fe-S-clusters) can form spontaneously *in vitro* [5] and that *in vitro* reconstitution[1] of proteins is constantly reported (e.g. [6]).

1.2.3 Transition metals - coordination geometry and ligand field splitting

Many of the biological metals are transition metals (see Figure 1.8), which have d-orbital electrons. The d-orbital electrons allow many of these metals to appear in a variety of oxidation states (compared to e.g. Mg and Ca, which is mainly limited to (0) and (II)). Many of the transition metals allow d-orbital hybridization in complex with ligands and thus coordination of more ligands and a higher variety of coordination geometries [5]. Common geometries of ligand-metal complexes with 3 and 4, and 5 and 6 ligands are shown in Figures 1.10 and 1.11, respectively.
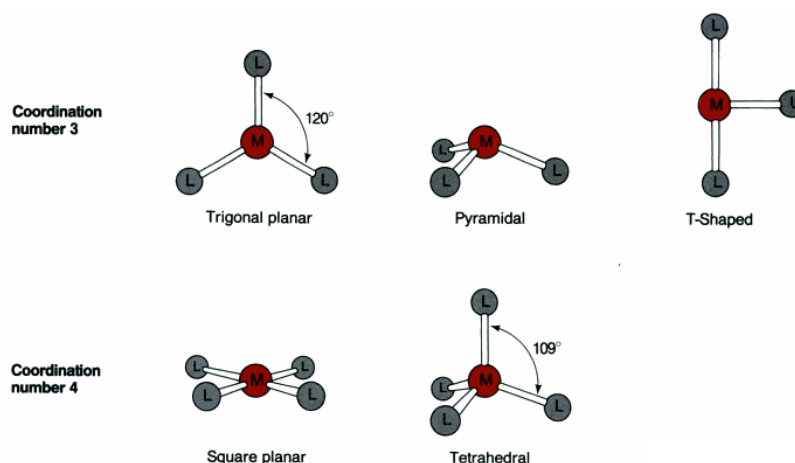
Figure 1.10: Common coordination geometries of metal-ligand complexes with 3 and 4 ligands. Figure reproduced from [5].

---

[1] Formation of the folded, metal-bound form of a protein from the unfolded polypeptide chain.
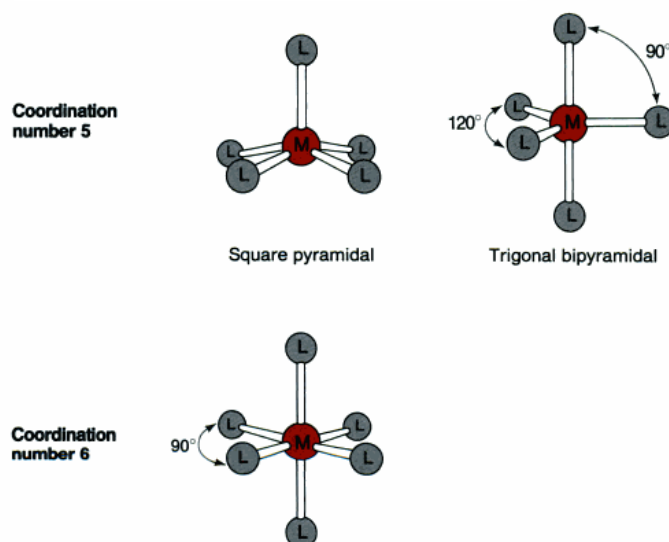
Figure 1.11: Common coordination geometries of metal-ligand complexes with 5 and 6 ligands. Figure modified from [5]

For a free metal ion, the energy levels of the d-orbitals are degenerate (equal). When a transition metal binds ligands, the d-orbitals change into lower- and higher-energy level orbitals. This is referred to as ligand-field splitting. The energy gap between the low- and high-energy orbitals is determined by both the nature of the ligands and the geometry of the ligand-metal complex. Depending on the geometry of the ligand-metal complex and the number of d-orbital electrons, ligand-field splitting leads to, or decides the number of, unpaired electrons (Figure 1.12) [5]. Unpaired electrons are responsible for paragmagnetism, a phenomenon with important consequences for various spectroscopical methods (see section 1.3) [5].
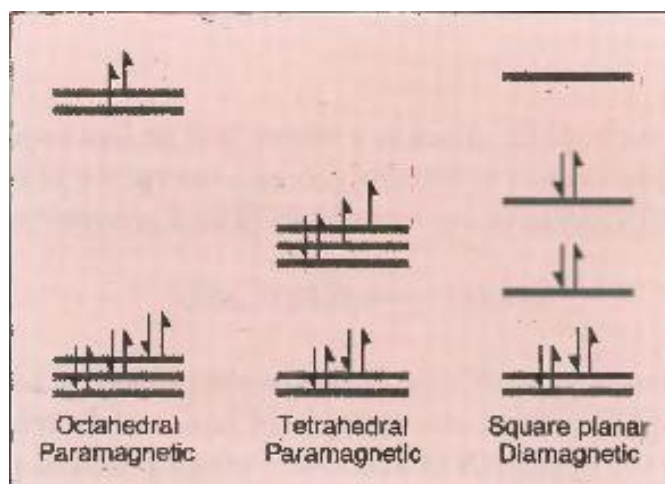
Figur 1.12: Ligand-field splitting diagrams. Ligand-field splitting diagrams shown for d^8 Ni(II) complexes with different coordination geometries. Energy levels and number of unpaired electrons depend on the coordination geometry. Figure reproduced from [5].

1.2.4 Structural zinc-binding sites

As already mentioned, many proteins have metal-binding sites whose function may be to facilitate the right tertiary structure of the protein. The majority of such sites are calcium- or zinc-binding sites [5]. The most common zinc binding sites are called "zinc fingers", which are small folded motifs that stabilize the structure of a protein. In these motifs the metal ion is bound to the protein by cysteins or histidine side chains in tetrahedral geometry [7]. Models of zinc-fingers and other structural zinc-sites are shown in Figure 1.13.
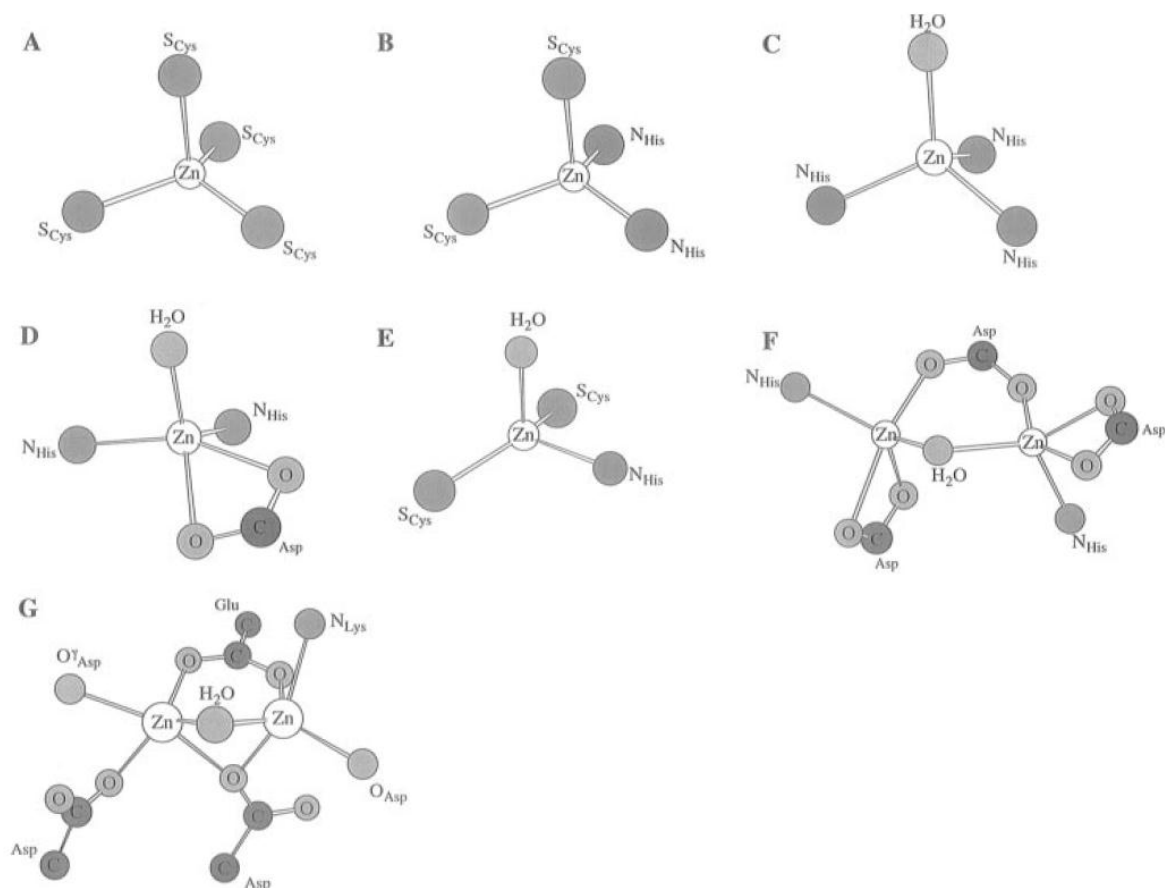
Figure 1.13: Structural zinc-binding sites. Schematic illustration of various structural zinc-binding sites, showing the zinc atom, the atoms to which it is bound (ligands) and the spatial relationship. The amino acids or molecules to which the coordinated atoms belong are denoted beside the atoms. Figure reproduced from [4].

1.2.5 Metal substitution in zinc-binding sites

For some proteins with metal-binding sites, the native metal ion can be substituted with other metal ions. This substitution can be achieved *in vivo* by restricting the growth medium of the native metal and supplying a non-native metal in excess (e.g. [8, 9]) or *in vitro* by reconstitution (e.g. [10, 11]). Both structural and catalytic zinc-binding sites have been shown to be metal-substitutable, and some enzymes have been shown to be catalytically active after both Co(II)-, Cu(II)-, Ni(II)- and Hg(II)-substitution. For many enzymes with a

catalytic zinc-site, Co(II)-substitution has produced enzymes with similar or even higher enzymatic activity [5].

1.2.6 Cobalt-containing proteins

Vitamine B_{12} is a cobalt-containing cofactor which is widely used in nature (e.g. in the enzyme methylmalonyl-CoA muthase [2]), and the majority of cobalt-containing proteins are vitamin B_{12}-dependent enzymes [3]. In fact, cobalt-containing non-vitamin B_{12} proteins (also called non-corrin cobalt proteins) are represented by only nine enzymes, all in which cobalt is found in five- or six-coordinated complexes with octahedral/square pyramidal geometry [12, 13]. Of these, only one is a common enzyme also found in humans (methionine aminopeptidase) while the other appear to be specific for certain strains of bacteria, fungi, yeast or algaea [12], and it has been postulated that non-corrin cobalt proteins are evolutionary relics from the time on earth when zinc was not readily available [3].

1.2.7 Metal transport and metalloproteins in bacteria

As already mentioned, affinity of a divalent metal ion for ligands in a polypeptide follows the Ivring-Williams series. Hypothetically, this means that if a polypeptide chain is surrounded by all metals of this series, all proteins should be copper-proteins. This provides a challenge for the cell in inserting the correct metal ion in a newly formed polypeptide chain. In bacteria such as *Escherichia coli*, this is achieved by separating the different metals into the cytosol and periplasm[2] according to their place in the Irving-Williams series using a complex system of metallosensory-, influx- and efflux-proteins. Metals like zinc and copper, which represent the metals with highest ligand affinity, are held in higher concentrations in the periplasm and very low concentration in the cytosol. The system relies on the metallosensory proteins to regulate complementary influx/efflux-proteins, both which are specific towards one metal and whose sensitivity towards activation corresponds to the metals place in the Irving-Williams series (Figure 1.14) [14]. The uptake of the correct metal by the individual polypeptide chains is thus achieved by the gradient of the metals that results from this system, directing certain polypeptides to the periplasm and by metallochaperones[3]. Nickel, required by a very limited number of enzymes, appears not to be released into the cytoplasm at all and is aqcuired at the transporter protein by metallochaperones and delivered directly to the nickel-requiring enzymes [14].

---

[2] Space between the cytosolic membrane and cell membrane.

[3] Chaperones are proteins that assist the correct folding of a newly formed polypeptide. Metallochaperones assist in delivering the correct metal ion to different polypeptides.
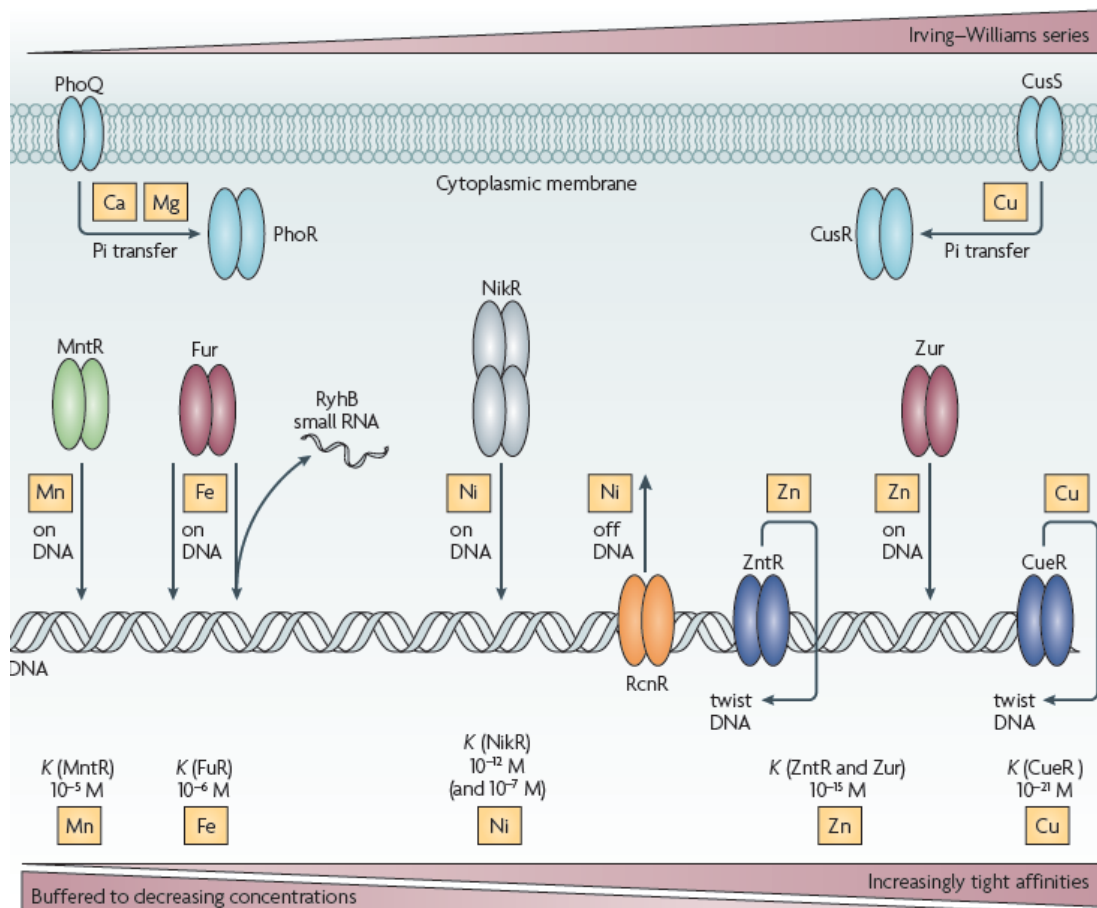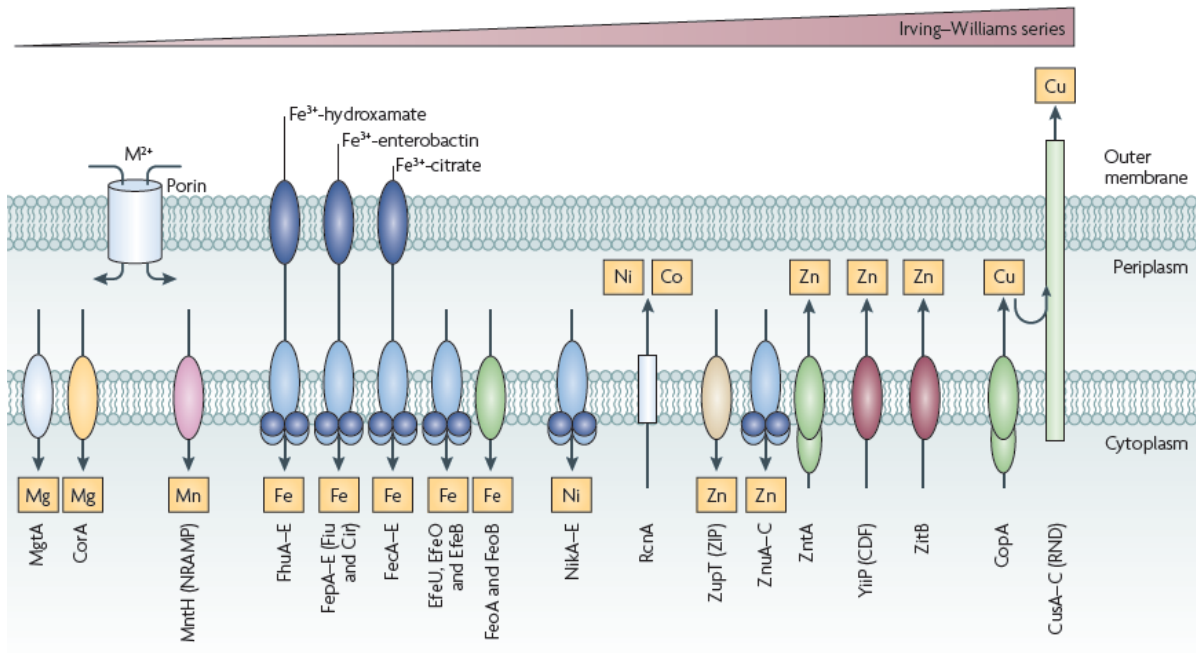
Figure 1.14: Regulation of cytosolic and periplasmic metal concentration in E. Coli. Metals of high ligand affinity are transported into the periplasm, while metals with lower affinity are kept in the cytosol (top). The different transporter proteins are regulated by complementary sensory proteins that show a metal sensitivity relative to each other that corresponds to the Irving-Williams series. Not shown is the cobalt-metallosensory protein, RcnR, which has a dissociation constant of 2×10^{-7} [15]. Figure reproduced from [14].

1.3 Methods in structural proteomics

Much of the investigations in structural proteomics are based on different spectroscopic techniques. Common for all spectroscopic methods is exposing a sample to electromagnetic irradiation of a given frequency or frequency range, upon which scattering or absorbance of this irradiation by the sample leads to a change in the recorded output intensity [5]. Different spectroscopic techniques operate over different and limited frequency ranges within this broad spectrum (Figure 1.15[4]).
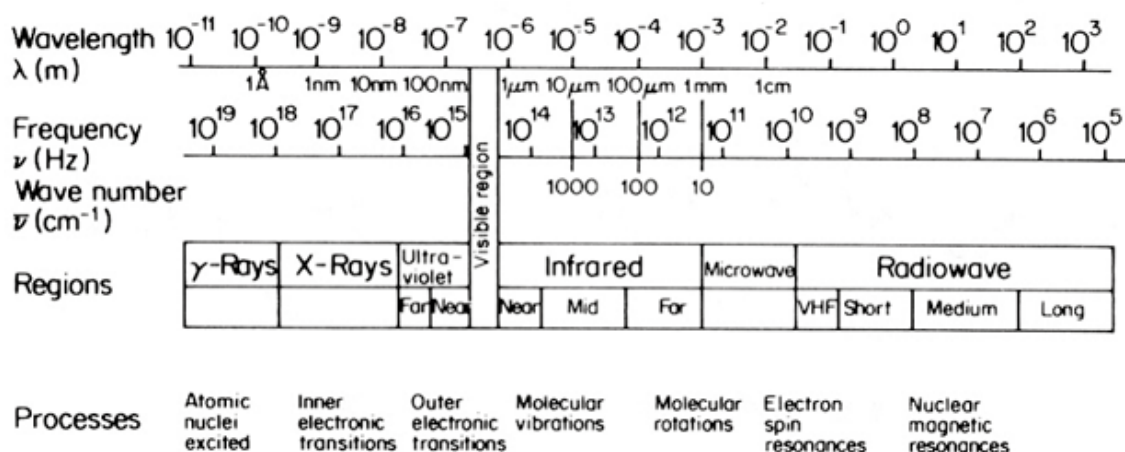
Figure 1.15: The electromagnetic radiation spectrum. The different frequency ranges of the electromagnetic radiation spectrum with subclassification of frequency range. Molecular, nucleic or electronic influence of the different frequency ranges are also indicated.

1.3.1 Nuclear magnetic resonance (NMR) spectroscopy

X-ray crystallography and NMR spectroscopy are the only two experimental techniques providing complete structural elucidation of a biomolecule. X-ray crystallography is older, better developed and still the most frequently used method. However, nuclear magnetic resonance (NMR) spectroscopy has also become a very important tool for structural biology and –biochemistry. Below, the basic principles of NMR spectroscopy and the experiments most relevant to this thesis are described.

1.3.1.1 Theoretical basis for NMR and NMR spectroscopy

This section is adapted from [16].

NMR spectroscopy is based on different absorption characteristics for different nuclei when a molecule is placed in a magnetic field and irradiated with electromagnetic radiation in the radio frequency region. The absorption of electromagnetic waves depends on magnetic nuclei (with I≠0), characterized by a magnetic dipole moment arising the nuclear spin. The nuclear magnetic moment by its spin quantum number, I, which can be half or whole numbers (0, 1/2, 1, 3/2, etc.). A given isotope of an element always has the same spin

---

[4] Picture downloaded from http://www.science4heritage.org/survenir/approach/?c=NIR-spectroscopy.

number. A spin number of 1/2 gives readily obtained spectra for the corresponding nuclei, because they have a spherical charge distribution and unique energy transition. Examples of nuclei with a spin number of 1/2 are ^{1}H, ^{13}C and ^{15}N. These nuclei are the most frequently used in NMR spectroscopy of biomolecules.

When nuclei with a spin number of 1/2 are placed in a magnetic field two energy levels, α and β, take different energy values and energy degeneration is absolved (called Zeeman's effect). A slight majority of the nuclei will exhibit the lower energy state (α). With a stronger magnetic field the energy difference will be larger between the two energy states. This is expressed as

$$\Delta E = (h\gamma/2\pi)\, \boldsymbol{B}_0 \qquad\qquad (1)$$

where h is the Planck constant, γ the gyromagnetic ratio and $\boldsymbol{B}_0$ is the strength of the magnetic field. The nuclei can make transitions to the higher energy levels only when radiofrequency radiation with energy equal to ΔE is absorbed. This energy can also be described by the following equation:

$$\Delta E = h\nu \qquad\qquad (2)$$

where h is Planck's constant and ν is the frequency of radiation. The frequency required to induce transitions between the energy levels α and β for a given magnetic field strength can be determined from the equations 1 and 2, and is therefore given by

$$\nu = (\gamma/2\pi)\, \boldsymbol{B}_0 \qquad\qquad (3)$$

During absorption, the magnetic dipole moment of a nucleus deviates from its equilibrium position along the z-axis (the direction of the magnetic field) towards the xy-plane where it is precessing. This tipping of the magnetic dipole moment is recorded by an oscillator coil in the instrument producing a Free Induction Decay (FID) curve, which by a mathematical treatment called Furie transformation puts the frequency domain into correspondence with the time domain and is thus transformed into an NMR spectrum.

Because different nuclei are in different chemical environments the strength of the magnetic field experienced by the nuclei, and therefore also the frequency of radiation needed to induce transitions, will be different for different nuclei. Two methods are used to induce transitions in all spectra: continuous-wave(CW) and pulsed spectrometer. In CW, the whole radio frequency is irradiated in a continuous sweeping. This method is no longer used, especially for NMR of biomolecules. In the pulsed spectrometre, the sample is irradiated with a pulse of high power radiofrequency energy of all frequencies simultaneously, and is the most used method today.

1.3.1.2 NMR experiments with proteins

This section is based on material from the website Protein NMR - A Practical Guide by Dr. Vicky Higman, Department of Biochemistry , University of Oxford. All images are reproduced or modified from this webpage. Available at http://www.protein-nmr.org.uk/spectra.html (last visited 29.04.2012).

The natural isotopes of carbon and nitrogen (^{12}C and ^{14}N, respectively) are spectroscopically "silent" in NMR, meaning that they do not absorb radiation in NMR-experiments and thus prohibit coupling that some of the experiments depend on. Advanced NMR experiments with proteins therefore depend on proteins labeled with the NMR-detectable isotopes ^{13}C and/or ^{15}N. All experiments below depends on labeling with both isotopes unless otherwise stated. In the accompanying illustrations, atoms that are contribute to the respective spectra by cross peaks are colored pink, whereas atoms that are involved in transferring signals but are not represented with cross peaks are colored blue.

NMR-experiments for assignment of backbone atoms:

^{15}N HSCQ (2D – minimum labeling required: ^{15}N):
Shows cross peaks for each N-H^{N} pair of a protein. The spectrum is mostly represented by cross peaks from backbone N-H^{N}, but also from side chains of certain amino acids. Since Pro residues do not have N-H^{N} pairs, they are not visible in the spectra (Figure 1.16).
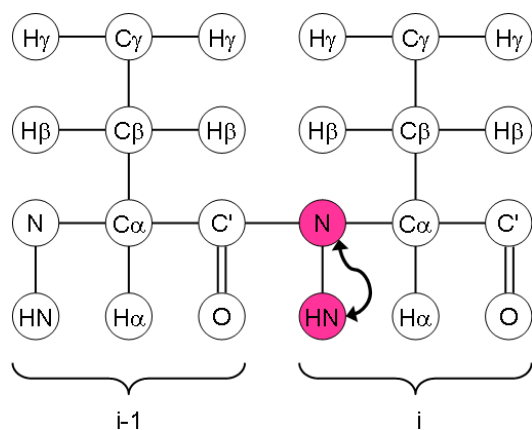
Figure 1.16: ^{15}N HSQC: Schematic representation of amino acids of a protein and the atoms involved in the effects recorded in an HSQC-experiment.

The ^{15}N HSQC is generally the first spectrum that is run on a new protein, and can give much information in itself. The two-dimensional N-H^{N} correlation is the basis for all the three-dimensional experiments described below. Because of this, the ^{15}N HSQC spectrum of a protein is usually overlaid to these spectra in order to give meaningful correlations.

HNCA (3D):

Correlates the N-H^N pair with C$^\alpha_i$ and C$^\alpha_{i-1}$, thus giving two cross peaks in the third dimension for each N-H^N pair (Figure 1.17). The cross peak for C$^\alpha_i$ is usually more intense.
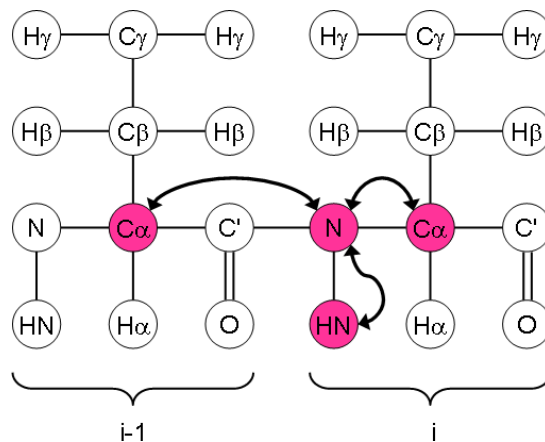
Figure 1.17: HNCA. Schematic representation of amino acids of a protein and the atoms involved in the effects recorded in an HNCA-experiment.

CBCANH and CBCA(CO)NH (3D):

CBCANH gives individual cross peaks for C$^i_\alpha$, C$^i_\beta$, C$^{i-1}_\alpha$ and C$^{i-1}_\beta$ for each N-H^N pair. CBCA(CO)NH gives individual cross peaks only for C$^{i-1}_\alpha$ and C$^{i-1}_\beta$ (see Fig. 1.26). Comparison of the two spectra makes it possible to distinguish which cross peaks belong to which amino acid (Figure 1.18).
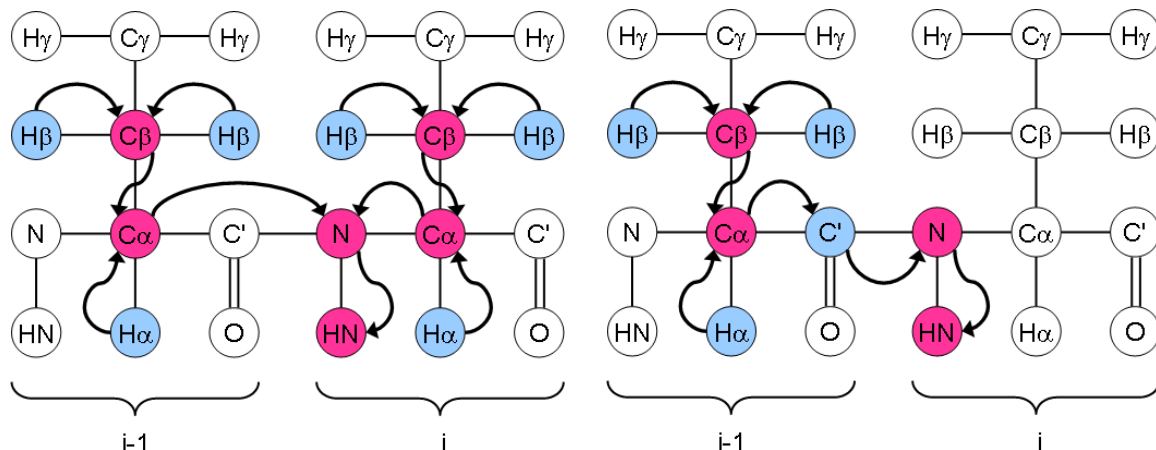
Figure 1.18: CBCANH and CBCA(CO)NH. Schematic representation of amino acids of a protein and the atoms involved in the effects recorded in CBCANH- and CBCA(CO)NH-experiments (left and right, respectively).

HN(CA)CO and HNCO (3D):

HN(CA)CO gives individual cross peaks for CO^i and CO^{i-1} for each N- H^N pair. HNCO gives individual cross peaks only for CO^{i-1} (Figure 1.19). Comparison of the two spectra makes it possible to distinguish which cross peaks belong to which amino acid.
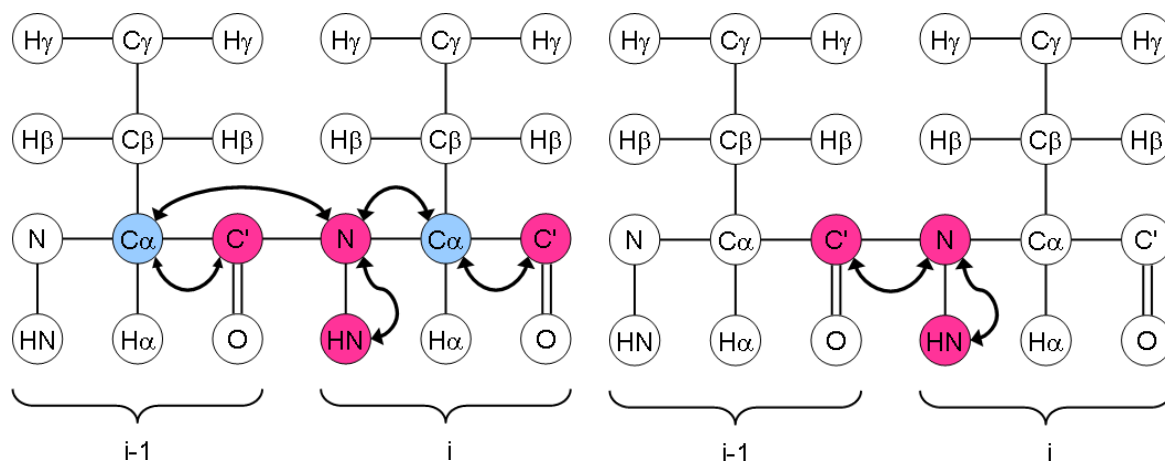
Figure 1.19: HN(CA)CO and HNCO. Schematic representation of amino acids of a protein and the atoms involved in the effects recorded in HN(CA)CO- and HNCO-experiments (left and right, respectively).

HBHANH and HAHB(CO)NH (3D):

HBHANH gives individual cross peaks for H_α^i, H_β^i, H_α^{i-1} and H_β^{i-1} for each N- H^N pair. HAHB(CO)NH gives individual cross peaks only for H_α^{i-1} and H_β^{i-1} (Figure 1.20). Comparison of the two spectra makes it possible to distinguish which cross peaks belong to which amino acid.
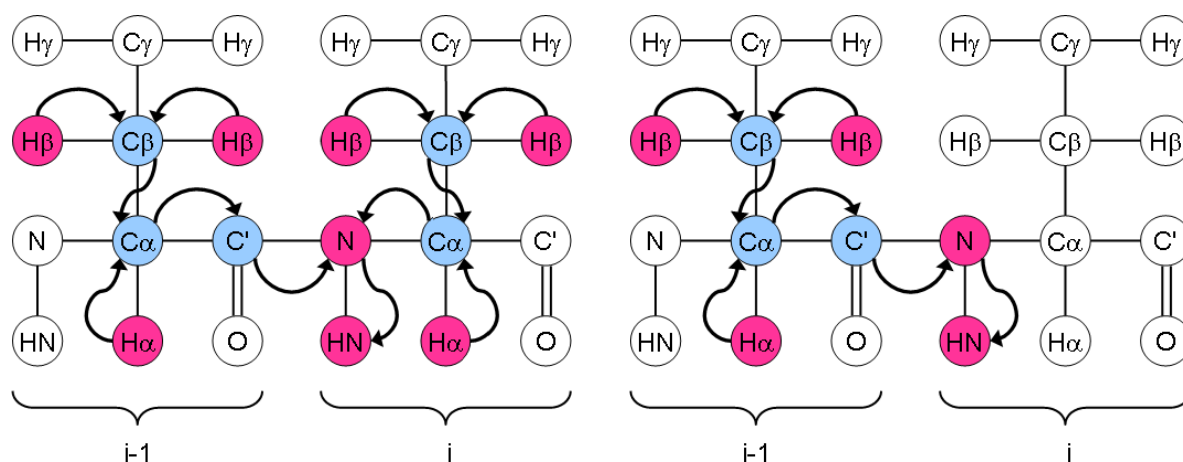
Figure 1.20: HBHANH and HAHB(CO)NH. Schematic representation of amino acids of a protein and the atoms involved in the effects recorded in HBHANH- and HAHB(CO)NH-experiments (left and right, respectively).

Spectra for assignment of side chain atoms:

HCCH-TOCSY (2D minimum labeling required: ^{13}C):

HCCH-TOCSY correlates and gives cross peaks for all carbon atoms of a given amino acid (except CO) and all hydrogen atoms of the same amino acid (except H^N) (Fig. 1.21).
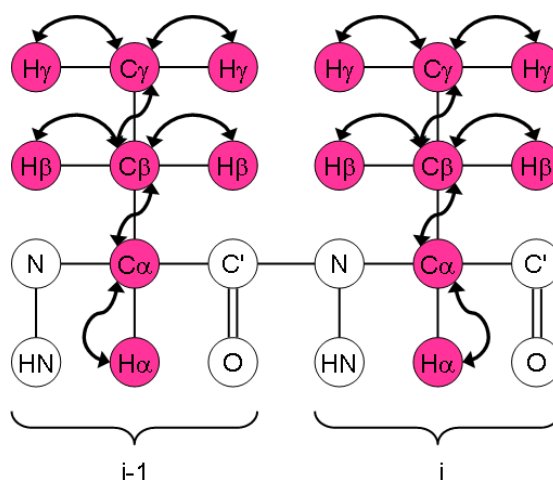
Figure 1.21: HCCH-TOCSY. Schematic representation of amino acids of a protein and the atoms involved in the effects recorded in a TOCSY-experiment.

1.3.1.3 NMR of proteins with paramagnetic metal ions

Paramagnetic metal ions (pMes) contain at least one unpaired electron. The unpaired electron(s) interact(s) with the nuclei of the atoms directly linked to the metal ion (contact interaction) and atoms within a radius of 10-30 Å (dipolar interaction) depending on the pMe [17]. Examples of pMes found naturally in proteins are Fe(II), Fe(III), Cu(II), Mn(II), Mn(III), Ni(II) and Co(II) [18-20]. pMes that are not native to a protein may also be introduced by in vitro biochemical/chemical methods or by creating recombinant proteins with pMe binding sites [21].

A paramagnetic metal ion influences both the chemical shift of the surrounding nuclei as well as their relaxation properties. As a result, signals from paramagnetically influenced nuclei are both shifted outside the diamagnetic region and are broadened. In fact, the paramagnetically influenced signals may in some cases be shifted by several hundred ppm, and their signal linewidth may be several hundred hertz (Hz), while the corresponding values for usual diamagnetic signals spans from -1 to 11 ppm and 2 to 15 Hz, respectively. Characterization of proteins with pMEs is therefore significantly more difficult and challenging compared to proteins with diamagnetic or no metal ions.

An example of a 1D ^{1}H NMR spectrum of a protein with a pME is shown in Figure 1.22.
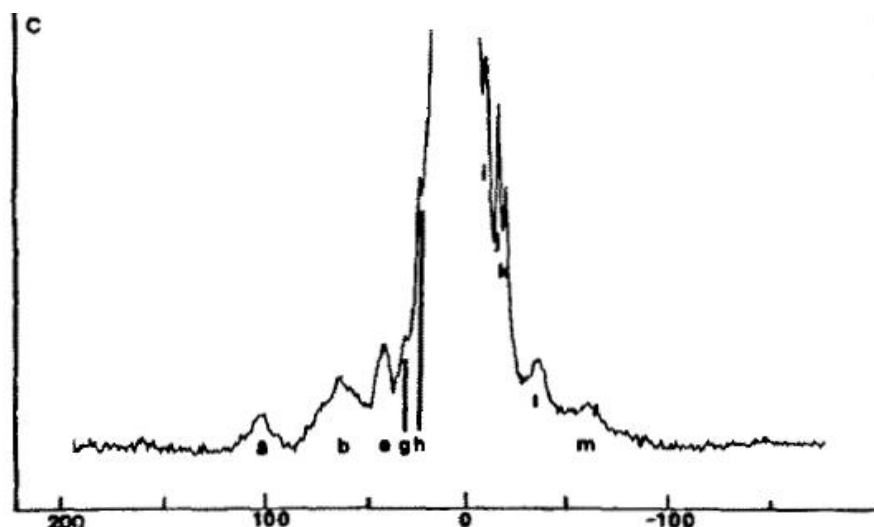
Figure 1.22: Paramagnetic ^{1}H NMR spectrum. ^{1}H NMR spectrum of a protein with a paramagnetic metal center, showing both high chemical shifts and signal broadening. Picture modified from [22].

In a protein with a pMe, the observed chemical shift for a given nucleus can be expressed as

$$\delta^{obs} = \delta^{dia} + \delta^{pm} \tag{4}$$

where δ^{obs} is the observed chemical shift (in ppm), and δ^{dia} and δ^{pm} is the diamagnetic and paramagnetic contribution to the observed chemical shift, respectively. Since the paramagnetic contribution exists of both contact- and pseudocontact shift contribution, (4) can be expressed as

$$\delta^{obs} = \delta^{dia} + \delta^{con} + \delta^{PCS} \tag{5}$$

where δ^{con} (contact contribution) is the chemical shift (also called contact shift) that results from unpaired electrons distributed through through atomic orbitals (chemical bonds) with the pMe, δ^{PCS} (pseudocontact contribution) is the chemical shift (also called pseudocontact shift) which results from interaction between the unpaired electron of the pMe and the nuclei through space (in effect an extra magnetic field). In a protein where a diamagnetic metal ion has been replaced by a pMe and little or no structural changes has occurred δ^{dia}, δ^{con} and δ^{PCS} constitute the largest contributions [21]. However, the contact shift (δ^{con}) does not influence atoms separated from the pMe by more than four σ-bonds [17] or 7 Å (for pMes with 3d-electrons) [21], and for residues that are not directly bound to or adjacent to residues bound to the pMe the contribution to observed chemical shifts are δ^{dia} and δ^{PCS}.

δ^{dia}, the chemical shift which a nucleus would have if the protein did not contain a pME, is unaffected by a pMe, and for a protein where the structure of the diamagnetic form is solved by NMR it is already known. From a given NMR spectrum of the paramagnetic form, the δ^{PCS} can then be calculated using equation 5 for all nuclei which are not directly bound to the pME and thus experience no contact contribution:

$$\delta^{PCS} = \delta^{OBS} - \delta^{dia} \tag{6}$$

As previously mentioned, the δ^{PCS} results from the magnetic field created by the unpaired electron of the pMe, and can be thought of as perturbations of the δ^{dia} in residues which do not experience contact shift contribution (Figure 1.23).

Figure 1.23: HSQC of a protein with dia- and paramagnetic metal ion. Schematic representation of the δ^{PCS} contribution to the chemical shift in an HSQC experiment. The black circles represent the chemical shift before replacement of the diamagnetic metal ion by a pMe, and the gray circles represent the chemical shift in the same protein with a pMe. Figure modified from [21].
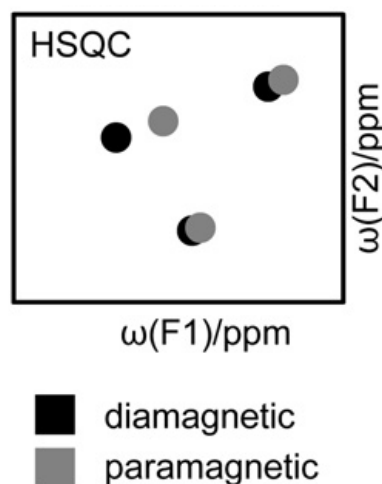
The magnetic field may contain components of both isotropic (equal in all directions) and anisotropic (direction dependent) magnetic susceptibility. Practically, this means that the magnitude of δ^{PCS} depends solely on the distance to the pMe in the areas surrounding the pMe where isotropic magnetic susceptibility dominates (thus a globular magnetic field), and more complicated in areas with anisotropic magnetic susceptibility (non-globular) [21]. For a protein in which the structure has been determined (or a preliminary structure exists), the observed PCSs can be used to create a mathematical model for the magnetic susceptibility which is called the magnetic susceptibility tensor. From this tensor, PCSs can be calculated for the nuclei (normally ^{1}H, as it is most sensitive to the magnetic susceptibility) of residues that have been assigned or may be used to predict PCSs). A high degree of agreement between observed and calculated PCSs indicates that a structure is a good representation of the protein conformation [23, 24].

NOE-experiments are the mostly used tool for investigating the secondary and tertiary structure of a protein, giving information about proximity of atoms in different residues. This information can be used to limit the number of possible conformations of a polypeptide chain. However, NOE is unsuitable for certain types of structural motifs [21] and is limited to the relatively short hydrogen bonds [23]. Since δ^{pcs} is proportional to $1/r^3$, where r is the distance between paramagnetic metal ion and certain proton, the PCS thus in itself contains geometric information that may be exploited as structural constraints in structure calculation or in refinement of an already determined structure. The method has been used for both complete determination of protein structure and for refinements to already solved structures [21]. NMR of paramagnetic metalloproteins is therefore used as an alternative to NOE, and represents an exciting field within NMR spectroscopy.

1.3.2 UV-Visible spectroscopy

1.3.2.1 Theoretical basis

A wide range of biomolecules absorb light in the ultraviolet and visible frequency range (see Figure 1.14), often at characteristic wavelengths. As UV-visible light of a given intensity passes through a sample, a fraction of this light is absorbed by the sample. This fraction is proportional to the concentration of the species responsible for absorption (c), the lights path length through the sample (l) and the molar extinction coefficient (ε) [2]. This relationship, called the Lambert-Beer law, is the basis of UV-vis spectroscopy and is expressed as

$$\log (I_0/I) = \varepsilon cb \qquad (7)$$

where I_0 is the intensity of the irradiating light, I is the intensity of the light leaving the sample. $\log (I_0/I)$ is called the absorbance, and is often simplified by the designation I (or *A*). The molar extinction coefficient varies with the absorbing species, the wavelength of irradiating light and solvent [2].

1.3.2.2 Cobalt(II) as UV-vis spectroscopic probe in metal-binding sites

Cobalt(II) can sometimes be inserted into metal binding sites of proteins as a substitute for the native metal, upon which coordination of ligands leads to a splitting of the energy levels of the d-orbital electrons. When irradiated with light the Co(II)-ligand complex absorbs light at specific wavelengths owing to excitation and relaxation of the d-orbital electros, also known as d-d-transitions [26]. The wavelengths at which this absorption occurs depends on the nature and number of the ligands as well as the cobalt-ligand coordination geometry [25]. UV-Vis absorption spectra of tetrahedral four-coordination of Co(II) in zinc-binding sites ("zinc-fingers") with different numbers of Cys- and His-ligands are shown in Figure 1.24.

Figure 1.24: UV-Vis absorption spectra for Co(II)-coordination by different zinc-finger motifs. Typical UV-Vis absorption spectra in the 500-800 nm range is shown for tetrahedral four-coordination of Co(II) by zinc-binding sites with four Cys (dotted line), three Cys and one His (broken line) and two Cys and two His (solid line). Figure reproduced from [25].
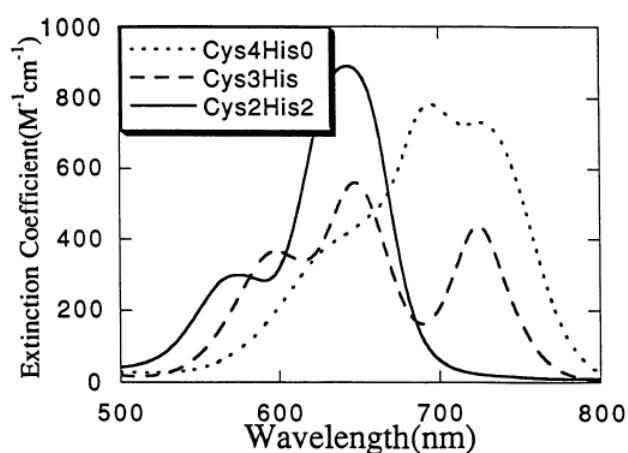
1.4 Biology of oxidative stress

1.4.1 Oxidative stress – underlying mechanism and biological significance

Molecular oxygen (O_2) is crucial to all animals, yet its utilization poses some biological challenges. In particular, aerobic respiration inevitably produces compounds called reactive oxygen species (ROS) as a byproduct. These molecules and free radicals may oxidize and/or induce radical chain reactions and thereby damage cellular components. Cells use redox-coupled enzymes to scavenge ROS and prevent damage. The imbalance of high generation of ROS and relatively low capacity of the redox-defense systems leads to harming processes generally referred to as oxidative stress [27].

In the process of oxidative phosphorylation, electrons flow through a series of enzyme complexes (the electron transport chain, ETC) in the mitochondria to indirectly power the production of ATP. At the end of this chain, O_2 is needed to accept electrons by reduction to water (H_2O). In the course of the electron transport, however, single electrons may leak from the radical-intermediate forms of the enzyme complexes to free O_2 and thus produce a free radical, superoxide (O_2^-), the main endogenous source of ROS (Figure 1.25).
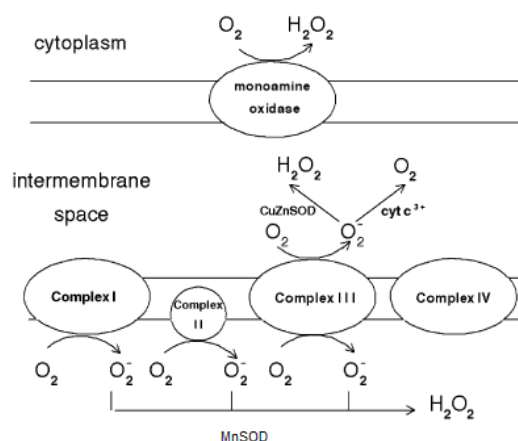
Figure 1.25: Formation of superoxide in the electron transport chain. Figure modified from [28].

The formation of O_2^- is expected to vary greatly between different tissues, but has been estimated *in vitro* to constitute 1-2 % of the total oxygen consumed [28, 29]. When formed, O_2^- may be converted to hydrogen peroxide (H_2O_2) by superoxide dismutase (SOD) in the mitochondrial matrix (Mn-SOD) or cytosol (CuZnSOD) (see Fig. 1.23). O_2^- can also spontaneously donate its singlet electron back to the ETC, thus restoring O_2 [28]. H_2O_2 is reduced to H_2O by either glutathione peroxidases (GPx), peroxiredoxins (Prx) or catalases (CAT) (Figure 1.26) [30].
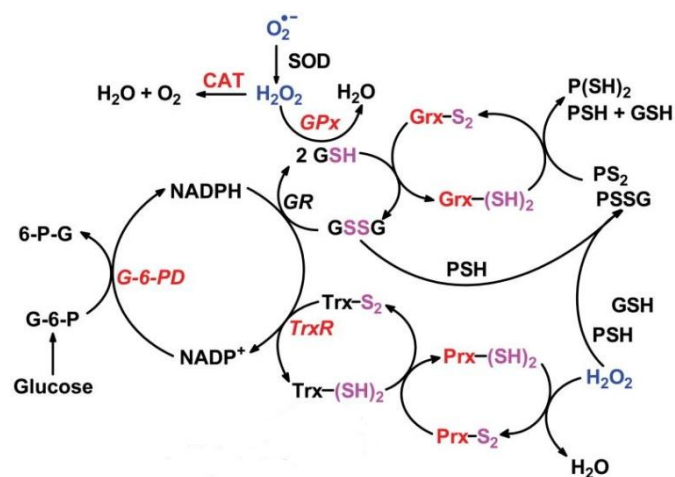
Figure 1.26: ROS scavenging and redox coupled enzymes. Overview of part of the redox-coupled enzymes and substrates involved directly and indirectly in scavenging O_2^- and H_2O_2. SS or S_2 indicates the oxidized form of an enzyme and SH the reduced form. Figure modified from [30].

When superoxide or hydrogen peroxide escape the scavenging mechanisms, they may attack and damage cellular components directly, or be converted to even more reactive ROS or reactive nitrogen species (RNS). ROS- and RNS-attack on DNA represents a major part of spontaneous mutations, and thus represents the most damaging effects of oxidative stress. Attack by ROS and RNS on polyunsaturated fatty acids produces lipid peroxyl radicals and lipid hydroperoxides, which in turn leads to free radical chain reactions with neighbouring lipids and can ultimately lead to harmful lipid derivatives or attack on DNA or protein. [29] ROS and RNS may directly attack both the backbone and sidechains of proteins, and can thus lead to modified or lost enzyme activity. Depending on the protein, this may lead to a temporary stimulation or inhibition of metabolic pathways, gene expression, signal transduction, ion influx or efflux, etc. [29]. Accumulation of oxidized protein is associated with various age-related diseases [31].

1.4.2 Oxidative stress and methionine

Methionine (Met) is in many aspects an especially important amino acid. It is used as the first amino acid during translation from RNA to protein, thereby initializing protein synthesis, and can therefore be a rate determining factor in translation when a cell is low in Met. In humans, Met is an essential amino acid, which means that it cannot be synthesised *de novo* or from other amino acids [32].

Met residues are particularly susceptible to oxidations, even by weak ROS, producing methionine sulfoxide (Met-SO) [31]. The oxidation of Met by ROS is non-enzymatic, and since the sulfur-atom of the Met sidechain is a prochiral centre[5] it therefore produces two

---

[5] A prochiral centre is an atom of a molecule which is non-chiral, but by chemical reaction (and thus desymmetrization) can give rise to a chiral centre. (Explanation adapted from IUPAC Gold Book, online at http://goldbook.iupac.org/P04859.html, last visited at 23.03.2012).

different diastereoisomers, namely Met-S-sulfoxide (Met-S-SO) and Met-R-sulfoxide (Met-R-SO). As far as is currently known, the oxidation produces equal amounts of both stereoisomers (i.e. a racemic mix) [32]. Depending on the position within a protein, oxidation of Met to Met-SO may impair the biological or catalytical function of the protein [33, 34] or leave it vulnerable to further oxidation. Also, simply degrading oxidized proteins is energetically inefficient because the protein then has to be resynthesized [35]. Enzymes to reduce Met-SO back to Met is therefore essential for an organism and is done by methionine sulfoxide reductases (Msrs) [34].

1.4.3 Methionine sulfoxide reductases

Enzymes that reduce Met-*S*-SO and Met-*R*-SO to Met are called MsrAs and MsrBs, respectively. An MsrA was first reported in 1979 and MsrB in 2002 [34], and most organisms have subsequently been demonstrated to exhibit both of these enzymes [35]. However, some prokaryotes lack MsrB, have a single enzyme to reduce both stereoisomers and/or have separate enzymes for reduction of free Met-*R*-SO [32, 35].

Cells keep a pool of free amino acids in the cytosol for peptide synthesis, and Met-SO thus exists both as free Met-SO (fMet-SO) and as part of a protein (pMet). [32] Msrs show a substrate preference for protein-bound Met-SO [34]. However, in mammals, MsrA has a somewhat low substrate specificity, and can reduce both peptide Met-*S*-SO, free Met-*S*-SO and probably non-amino acid sulfoxides. Mammalian MsrBs on the other hand, are specific towards reduction of Met-*R*-SO in peptides and reduce fMet-*R*-SO with very low activity (Figure 1.27) [32].
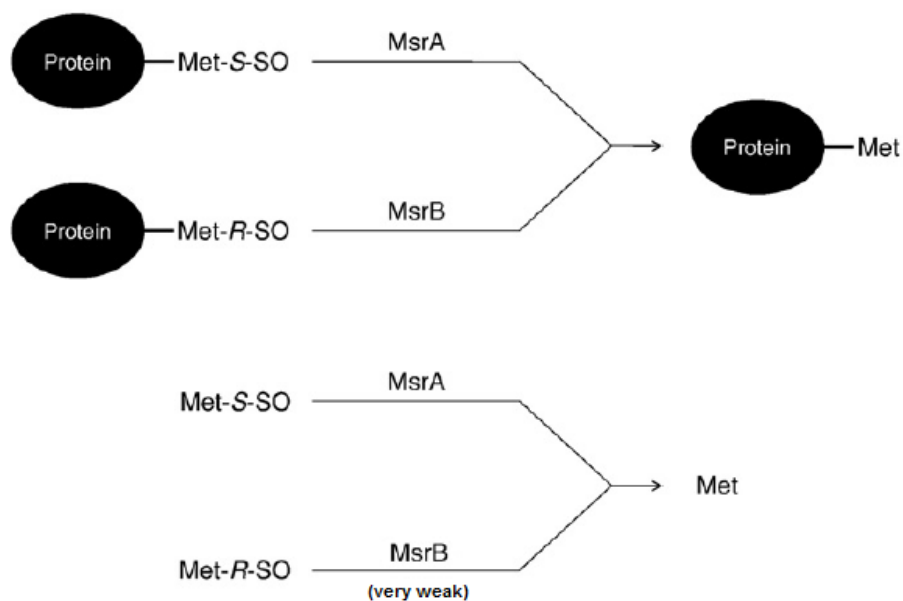
Figure 1.27: Met-SO to Met reduction in mammals. Figure modified from [32].

1.4.4 Intracellular distribution and structural difference of MsrA and MsrBs

In mammals, a single gene exists for MsrA, whereas MsrB is encoded by three different genes (MsrB1-3). In humans, MsrA is found only in the mitochondria, while in mouse and rat it has also been found in the cytosol and nucleus . MsrB1 is located in the cytosol and nucleus and MsrB2 only in the mitochondria. In humans, MsrB3 exist as two forms, MsrB3A and MsrB3B, the former being found in the endomplasmic reticulum (ER) and the latter in mitochondria. In mice, MsrB3 exists in a single form and is found in the ER (Figure 1.28) [33].
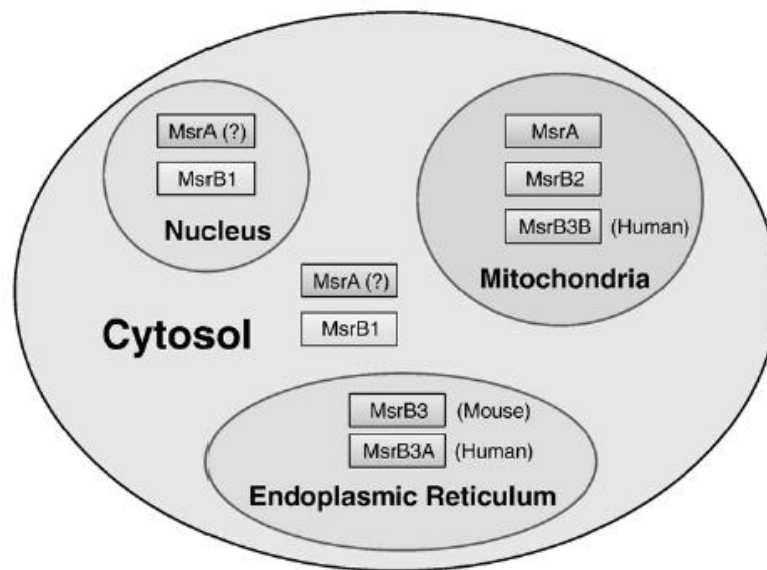
Figure 1.28: Distribution of MsrA and MsrBs in the cells of human and mouse. Figure reproduced from [33].

MsrA and MsrBs are targeted to the different cell compartments by different signal-peptides at the N- and C-terminus, and where one form is targeted to more than one location it is achieved by alternative splicing, thus yielding different mRNAs. The two forms of human MsrB3 are also a result of alternative splicing [33].

MsrA and MsrB lack homology in sequence and structure, yet carry out the same (but mirrored) chemical reaction and is thus an example of convergent evolution [35]. However, there is a high degree of homology in the individual genes between archaea, bacteria and eukaryotes, indicating that Msrs are ancient genes [35]. MsrA appears to represent a protein fold unshared by other protein families, while MsrBs show certain structural similarity with a protein family called Mss4 [36]. The structure of MsrB1 is described in section 1.4.7.

1.4.5 General catalytic mechanism and active site of Msrs

Met-*R*-SO and Met-*S*-SO have mirrored symmetry at the S-atom. This is reflected in the active sites of MsrA and MsrBs, illustrated by the three dimensional structure of the these enzymes in *N. meningitidis* and *N. gonorrhoeae*, respectively [34]. Although consisting of different amino acids, both active sites possess a hydrophobic and a hydrophilic pocket placed opposite relative to each other when compared with aligned substrates, and this arrangement is the basis for the stereoselectivity of these enzymes. The hydrophobic pocked stabilizes the end methyl- and methylene-group of the Met sidechain and the hydrophilic pocket participate provides a network of hydrogen bonds that activate the catalytic Cys and stabilize the transition state (Figure 1.29) [34, 37].
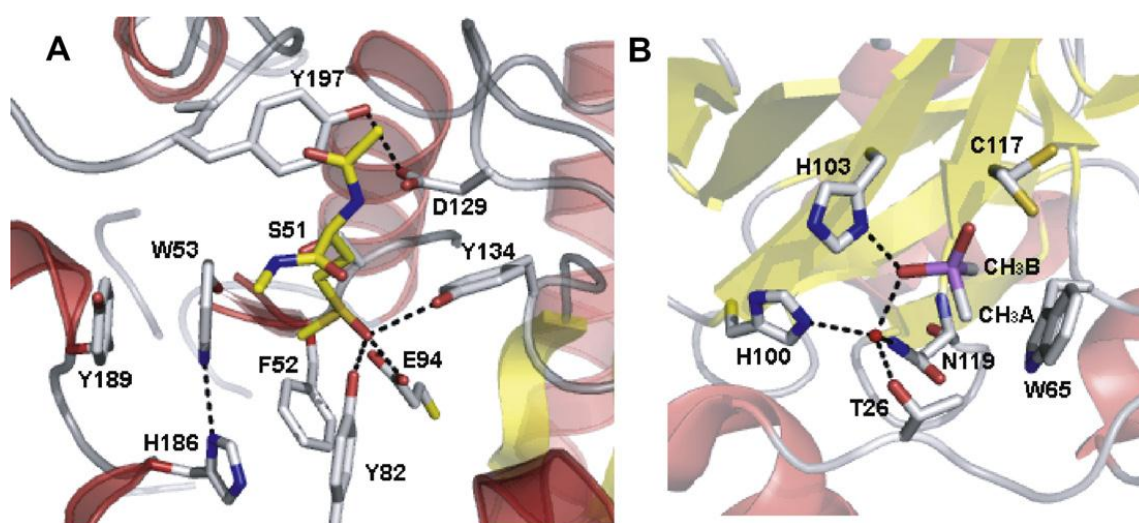
Figure 1.29: Active sites of MsrA and MsrB. The active sites of MsrA from *N. meningitidis* (A) and MsrB from *N. gonorrhoeae* (B) in complex with AcMetSONHMe and cacodylate, respectively. AcMetSONHMe mimics Met-*S*-SO and cacodylate mimics Met-*R*-SO, and the structures were found with X-ray crystallography. In (A), the hydrophobic pocket is formed by Phe-52 and Trp-53 in MsrA and the hydrophilic pocket is formed by Tyr-82, Glu-94, Tyr-134 and Asp-129. Ser-51 has replaced the catalytic Cys, Cys-51. In (B), the hydrophobic pocket is formed by by Trp-65 and Cys-117 and the hydrophilic pocket by Thr-26, Asn-119 and His-100 (via a coordinated water molecule) and His-103 [34]. Figure reproduced from [34].

Common for the reduction of Met-SO by Msrs is the following catalytic mechanism [34, 37]: 1) nucleophilic attack by a Cys (or Sec) side chain on the sulfur-atom of the sulfoxide, 2) formation of a sulfurane transition state substrate intermediate, 3) formation of Cys sulfenic acid (selenenic acid) and release of Met and 4) reduction of the Cys sulfenic acid to Cys (selenenic acid to Sec) directly or indirectly by Thioredoxin (Trx), restoring catalytic capability [38]. The mechanism of MsrB1 is shown in Figure 1.30.
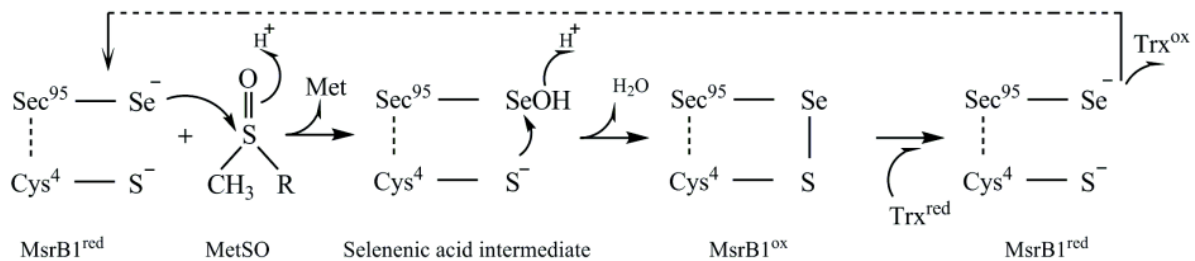
Figur 1.30: Catalytic mechanism of MsrB1. General catalytic mechanism of MsrB1. Residues involved directly in the catalytic mechanism are indicated together with the enzyme responsible for regeneration of catalytically competant protein. Figure reproduced from [38]

The Cys residue which attacks the sulfoxide is called the catalytic Cys (CC) or Sec. In mammalian MsrBs, step 4 is achieved slightly differently by the different Msrs. In MsrB1, the selenenic acid forms a selenylsulfide bond with a resolving cysteine (RC), before being reduced back to Sec by Trx. MsrB2 and MsrB3 lack RCs, and the Cys sulfenic acid intermediate is directly reduced by Trx (Figure 1.31) [33].
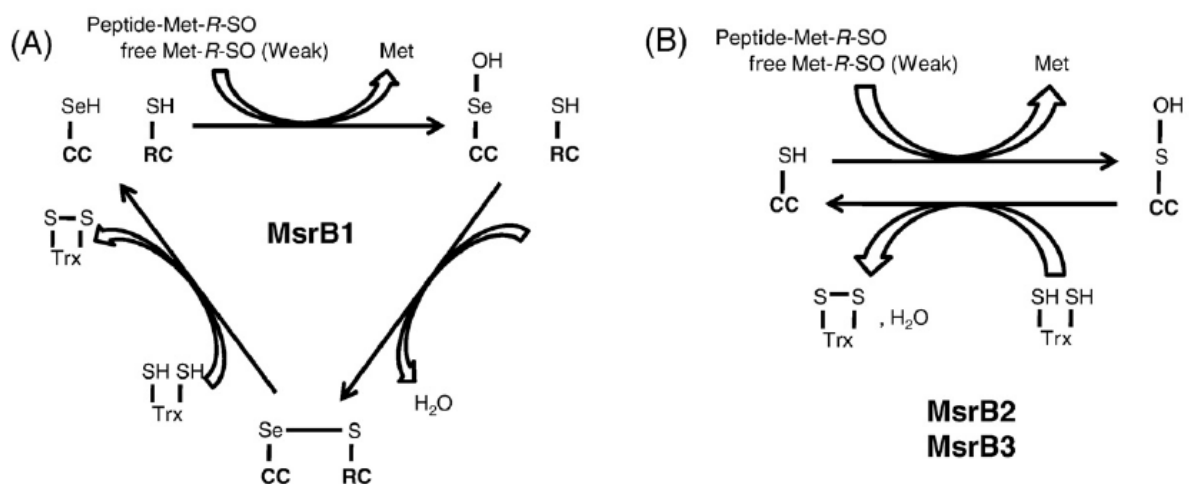
Figure 1.31: Catalytic mechanism of the different mammalian MsrBs. Figure displaying the reduction of Met-SO and paths of regeneration of the reduced catalytic Cys (CC) by (A) MsrB1 and (B) MsrB2 and MsrB3. Figure modified from [33].

1.4.6 Mammalian MsrB1 is a selenoprotein

As previously mentioned, the Msrs in different species show a high degree of homology in sequence and structure, thus enabling use of the bacterial enzymes as structural models for the conformations of residues of the active sites. The multiple alignment of MsrBs from different organisms show that the catalytic residues of the active site are conserved, with important exceptions in mammalian MsrB1 (Figure 1.32) [37]. In particular, His-77, Val-81, Cys-95 and Asn-97 have been replaced by Gly-77, Glu-81, Sec-95 and Phe-97, respectively. In MsrBs with Cys as catalytic residue, residues that create a network of hydrogen bonds with the oxygen of Met-*R*-SO are needed to increase the effective nucleophilicity of the catalytic

Figure 1.32: Sequence alignment of MsrBs. Alignment of amino acid sequences constituting the active sites of MsrBs from various organisms. The catalytic Cys/Sec is shown in red, the three other residues discussed in the text are shown in green and the remaining conserved residues are marked gray. Numbering of residues refers to the sequence of MsrB1 from *Mus Musculus*. Figure modified from [37].

Cys. In mammalian MsrB1 such residues are not needed due to the high nucleophilicity/low pKa of the catalytic Sec [37, 39]. This has been shown experimentally in that G77H, E81V and F97N mutants of mammalian MsrB1, both individually and combined, have significantly lower catalytic activity compared to the wild-type enzyme [37]. Conversely, H77G, V81E, N97F mutants of mammalian MsrB2 and MsrB3 showed lower or diminished catalytic activity, both individually and combined [37].

Selenocysteine (Sec) is an amino acid structurally analogous to cysteine, with selenium in the place of sulfur. In regards to electronegativity, oxidation states, atomic radius and other physicochemical properties, selenium and sulfur are very similar (for further discussion of this, see below) [39]. Selenoproteins are represented by a small number of protein families, and are found in many (but not all) of bacterial, archaeal and eukaryal species [39, 40], in total in about 25 % of all organisms [37]. The human genome contains 25 genes coding for selenoproteins and the murine genome 24 genes [39, 41]. In eukaryotes, Sec is encoded during translation by the codon UGA, normally a translation termination codon. The mechanism by which this is achieved is quite complex, requiring a Selenocysteine Insertion Sequence (SECIS) in the 3'UTR region of mRNA and several protein factors [40]. In most selenoproteins, Sec is a redox-catalytic residue, and Sec to Cys mutation typically lowers activity of the protein by 10-1000 fold [37, 39]. However, many organisms have a homologous protein with Cys instead of Sec, often with comparable enzyme acitivities. The biological advantage of and evolutionary pressure for a catalytic Sec is therefore a subject to debate because of the high cellular «costs» of Sec incorporation [39]. Mammalian MsrB1 is a selenoprotein and contains a catalytic Sec in the place of a catalytic Cys and has the highest catalytic activity of the Msrs [33, 42]. Reduction of oxidized MsrB1 by Trx, however, is slower than the reduction of the MsrB2 and MsrB3 [43].

1.4.7 Structure of MsrB1 (*Mus Musculus*)

The structure of MsrB1 (*Mus musculus*) consists of eight β-strands connected by loops, and no α-helices. The β-strands make up two antiparallel β-sheets with three and five β-strands, respectively, and the protein can therefore be characterized as a β-fold protein. It consists of 116 amino acids and a structural zinc(II)-ion. The zinc ion is tetrahedrally coordinated by four Cys-side chains (Cys23, Cys26, Cys71 and Cys74), thus representing a classic zinc-binding site (see section 1.2.4), and these Cys-residues are located in two CxxC-motifs, one in each β-sheet. The catalytic Sec is located on a β-strand of the largest β-sheet. The N-and C-termini are long, flexible chains, and the catalytic Cys is found on the N-terminus (Figure 1.33).

Figure 1.33: Structure of MsrB1: Top view (top) and side view (bottom) of the structure of MsrB1, with structural and catalytic residues are shown in magenta as sticks and the zinc-ion shown as a blue sphere. Structure represented as the the mean (average) conformer in the MsrB1 pdb-file (PDB ID 2KV1).

The first NMR-data for MsrB1 from *Mus musculus* were published in 2007 [44] and the first structure was published in 2010 [42]. The published structure contains the 20 different calculated conformers with lowest target function (best fulfillment of structural constraints), and Figure 1.34 shows the 20 conformers overlaid and indicate a structurally fixed core structure and many allowed conformations of the long N- and C-terminal chains.
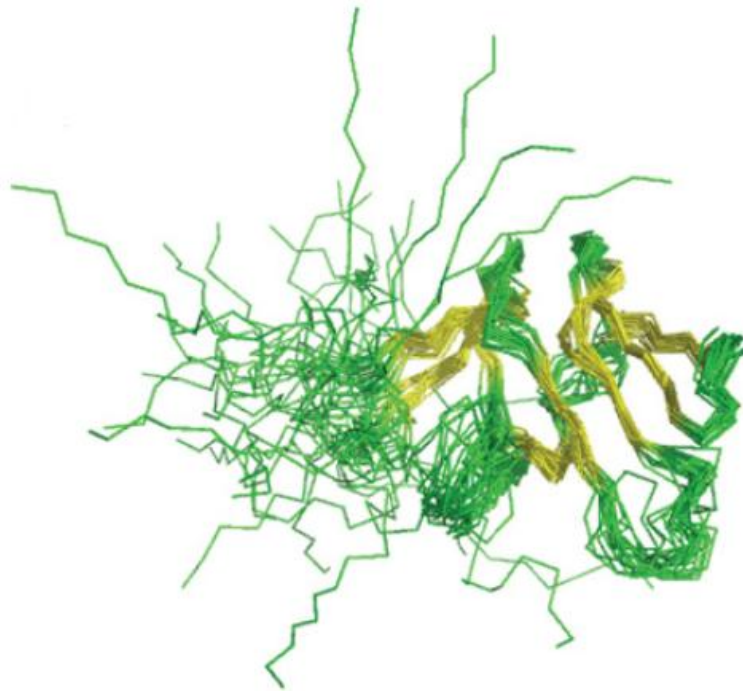
Figure 1.34: Calculated conformers of MsrB1: The 20 calculated conformers for MsrB1 with lowest target function overlayed, with β-strands indicated in yellow and loops and flexible chains in green. Figure modified from [42].

2 Aim of this study

This study is part of the ongoing research on the mammalian MsrB protein family. During the work of structural elucidation of MsrB1 from *Mus musculus*, paramagnetic signals were detected in the NMR spectrum of the protein when expressed in minimal medium supplemented with different metals. The metal responsible for these signals was identified as cobalt(II), and it was demonstrated that the His-tag of the protein was not responsible for metal uptake (Dobrovolska et al., unpublished results). The uptake of cobalt violates the Irving-Williams series, which places cobalt below zinc in ligand affinity. Together, these findings raise questions as to where cobalt is inserted in MsrB1, whether it has the same structure as the native protein and what causes the uptake of this metal.

The cobalt uptake, cobalt binding site and possible pH-titratable groups of Co-MsrB1 from *Mus musculus* will be characterized by UV-visible spectroscopy. This will be done by studying the UV-vis spectra of the purified protein expressed in different cobalt-supplemented growth media, and what changes of these spectra are imposed by pH. Further, metal removal will be attempted by exposing the protein to chelating agents. The concentration of Co-MsrB1 in purified protein samples from protein expressed in different growth media will be determined using the UV-vis absorption spectra. Backbone- and side-chain assignment of Co-MsrB1 ^{1}H, ^{13}C and ^{15}N signals of recorded NMR spectra will be attempted, and these assignments will be interpreted to gain structural information about the protein.

3 Materials and methods

3.1 Materials

3.1.1 Plasmids and host cell

The *Escherichia* coli strain ER2566 with a pET21b-MsrB1-Cys plasmid was available in the laboratory from previous research [42, 44], originally provided by Vadim N. Gladyshev based on the construct by Vadim N. Gladyshev and Hwa-Young Kim [37]. The plasmid construct contains 858 bp mutant cDNA coding for MsrB1U95C (MsrB1-Cys) from *Mus musculus* cloned into the *Nde*I/*Xho*I restriction sites of pET21b, resulting in a C-terminal His-tag (LEHHHHHH) added to the coding sequence. The amino acid sequence, element composition and predicted biochemical properties of His-tagged MsrB1U95C as predicted by ExPASy's ProtParam is shown in Appendix A. For convenience, the His-tagged MsrB1U95C recombinant protein will from here be referred to simply as MsrB1.

3.1.2 Growth media, buffers and solutions

All growth media, buffers and other solutions are described in Appendix B.

3.1.3 Other materials

Purification resin: Ni-NTA His Bind Resin (Qiagen). Consists of nitriloacetic acid with immobilized Ni^{2+} which has high affinity for poly-His like that of the His-tag of MsrB1. The resin binds MsrB1 during incubation, and releases MsrB1 upon addition of buffer with a high concentration of imidazole (elution buffer). This resin was used as described in the Qiagen Ni-NTA His-bind resin user manual[6]

SDS-PAGE-gel: NuPage® 4-12 % Bis-Tris gel, life technologies. Premade polyacrylamide SDS-PAGE gel cassettes. Protein samples were analyzed by SDS-PAGE as described in the product manual[7].

Dialysis membrane: Various experiments were conducted by dialyzing samples of MsrB1 against different buffers. For all these experiments, a cellulose membrane (Spectra/Por® Dialysis Membrane, Spectrum Laboratories, 6-8000 µm pore size) was first soaked in MQ water for minimum 30 minutes. The protein sample was then enclosed in the membrane by clamps, with a miminal amount of bubbles/air inside the dialysis membrane bag.

---

[6] Cell lysis and protein purification was performed as described in the Ni-NTA resin manual. Available online at http://www.qiagen.com/literature/render.aspx?id=128. Last visited 13.05.2012.
[7] NuPAGE® Technical Guide, life technologies. Available online at http://tools.invitrogen.com/content/sfs/manuals/nupage_tech_man.pdf. Last visited 13.05.2012.

3.1.4 Instruments

Ultrasonication: Branson Sonifier® 250 Analog, tapered microtip.

Measurement of protein concentration: All protein concentrations were measured using NanoDrop[TM] 1000 Spectrophotometer (Thermo Scientific) with the Protein A280 application version 3.7 (Thermo Scientific) as described in the product user manual[8].

pH-measurements: All pH-measurements of protein solutions were done using pHenomenal pH metre (VWR) with a 3 mm micro electrode (Schott Instruments).

UV-visible spectrometer: All UV-vis absorption spectra were recorded using quartz cuvettes (unknown manufacturer, 10 mm diameter, narrow chamber) in Ultrospec 2000 (Pharmacia Biotech) and the software application WaveScan[TM] (Pharmacia Biotech).

Mass spectrometer: All mass spectrometry-experiments were performed using a Bruker ultraflex III MALDI TOF/TOF mass-spectrometer.

NMR spectroscopy: All NMR experiments were recorded at 298 K on a Bruker Avance 600 MHz spectrometer equipped with a 5-mm z-gradient CP-TCI(H/C/N) cryoprobe at the NT faculty NMR center at NTNU, Gløshaugen and on an Avance 800-MHz equipped with a 5-mm z-gradient CP-TCI(H/C/N) cryoprobe at the Magnetic Resonance Centre (CERM) in Florence, Italy.

3.1.5 Software

Topspin: NMR spectra were processed using Bruker Topspin 3.1.

CARA: Computer Aided Resonance Assignment software version 1.8.4.2 was used for signal assignment of all NMR spectra.

PyMol: PyMol version 1.4.1 was used for visualization of protein structures from .pdb-files available at the webpage for Research Collaboratory for Structural Bioinformatics Protein Databank.

Molsoft ICM browser: Molsoft ICM-browser version 3.7-2b was used to visualize protein structures using the same files as for PyMol and for interatomic distance calculations.

Lucidchart: Lucidchart was used to create flowcharts of the experimental procedures. Available online at www.lucidchart.com.

---

[8] NanoDrop 1000 Spectrophotometer v3.7 User's Manual, Thermo Scientific. Available online at http://www.nanodrop.com/library/nd-1000-v3.7-users-manual-8.5x11.pdf. Last visited 13.05.2012.

3.2 Methods

3.2.1 Cell growth and production and purification of MsrB1

Precultures were made from a frozen 15 % glycerol stock-solution of *E. Coli* carrying pET-21b with the MsrB1 construct. Cells from the stock-solution were transferred to a sterile 100 mL flask with 25 mL sterilie LB medium, containing ampicillin (100 µg/ml), using a sterile platinum thread loop. The precultures were then incubated at 37 °C and 220 rpm for 8 hours or overnight. For culture growth in M9 medium, the LB preculture was first inoculated into a new M9 preculture.

8-10 mL of the preculture was then inoculated into a 2L flask containing 500 mL of sterile growth medium under sterile conditions. The culture was then incubated at 37 °C and 220 rpm under aerobic conditions and grown until the measured OD_{600} was 0.6-0.8 using sterile growth medium as reference. The culture flask was then cooled on ice for 5 minutes following addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG) to a total concentration of 1 mM to induce gene expression and production of MsrB1. The culture was then incubated for 37 °C and 220 rpm for 3.5-4 hours, followed by centrifugation at 7000 rpm for 5 minutes. The supernatant was removed, and the cell pellet weighed and frozen (- 20 °C).

The specific growth media and culture growth procedures for production of the different metal-containing forms of MsrB1 are described in section 3.2.1.1-3.2.1.4.

3.2.1.1 Zn-MsrB1

8-10 mL of an LB preculture was transferred to a 2L Erlenmeyer flask containing 500 mL sterile LB with ampicillin (100 µg/ml), and culture growth and gene expression was carried out as described in 3.2.1.

3.2.1.2 Co-MsrB1

Co-MsrB1 was produced by initially growing the bacteria in sterile M9 medium supplemented with $CoCl_2$ (M9-Co medium) and followed by gene expression as described in 3.1.5. Since presence of both Co and Zn in the growth medium will give both Co- and Zn-MsrB1, Zn had to be diminished from the precultures before inoculation into M9-Co medium to obtain only Co-MsrB1. As direct inoculation from the glycerol-stock into M9 medium rarely results in culture growth, 0.5 mL of an LB preculture was inoculated into 25 mL M9-Co medium in a 100 mL Erlenmeyer flask and incubated at 37 °C and 220 rpm overnight. Then, 0.5 mL of the M9-Co preculture was reinoculated in a series of 6-8 M9-Co precultures that were grown overnight. Finally, 8-10 mL of M9-Co preculture was inoculated into a 2 L Erlenmeyer flask containing 500 mL M9-Co medium and culture growth and gene expression was carried out as described in 3.2.1.

3.2.1.3 Zn- and Co-MsrB1

To study the ratio of the two different forms of MsrB1 when both metals were present during gene expression, the bacteria were grown in sterile M9 medium supplemented with 10 μM $CoCl_2$ and $ZnSO_4$, in LB supplemented with 50 μM $CoCl_2$ and LB supplemented with 10 μM $CoCl_2$. For the M9 medium, one preculture of the same medium was inoculated from an LB preculture like described in 3.1.2 before inoculation. For the LB media, inoculation was done from an LB preculture. Culture growth and gene expression was carried out as described in 3.2.1.

3.2.1.4 Ni-MsrB1

To see if MsrB1 would also bind nickel, the bacteria were grown in sterile LB medium supplemented with 10 μM $NiCl_2$ following inoculation from an LB preculture, and the culture grown as described in 3.2.1.

3.2.1.5 Effects of Co(II)-concentration on culture growth

As previously mentioned, M9 media containing 10 μM $CoCl_2$ had prior to this thesis been shown to lead to formation of Co-MsrB1. It was therefore important to determine if this Co(II)-concentration represented stress conditions to the *E. Coli* strain used in the experiments. If so, formation of Co-MsrB1 could possibly be attributed to a result of Co-stress.

To evaluate if 10 μM $CoCl_2$ was a stress condition, *E. Coli*-strain ER2566 was grown in different M9 media with a constant Zn-concentration (10 μM) and varying Co-concentration (2-150 μM), and the OD_{600} was monitored until it reached approx. 0.6 or as long as time would allow. To see how the concentration of Co(II) affected formation of Co-MsrB1, gene expression was induced in cultures that reached OD_{600} of 0.6.

3.2.2 Purification of MsrB1

Cell lysis and protein purification was performed as described in the Qiagen Ni-NTA His-bind resin user manual[9].

3.2.2.1 Cell lysis

The cell pellets were resuspended in 5 volumes of lysis buffer in 4 °C. The typical cell pellet from a 0,5 L culture varied from 1-3 g, depending on the growth medium. The cell suspension was ultrasonicated in a thick 20-30 mL glass tube in an ice bathe with the output control between 2.5-3 and duty cycle between 20-30%. The duration of sonication was approximately 10 minutes per 10 mL of cell suspension, with short breaks after 10 minutes continuous sonication for the bigger volumes. Samples of 5 µL of the cell suspension were analyzed quasi-quantitatively with Bradford Assay to assess the progress of the cell disruption.

3.2.2.2 Protein purification

After sonication, the lysate was clarified by centrifugation at 14000 rpm for 45 minutes. The supernatant was filtered using a plastic syringe with 0.45 µm filter. Purification of MsrB1 was based on liquid affinity chromatography by the use of Ni-NTA resin. Approximately 0.7 ml of Ni-NTA resin per liter culture were equilibrated in binding buffer using follow procedure: 1.5 mL of Ni-NTA resin (50% ethanol) was mixed with 10 mL of binding buffer and mixed for 5 min on a shaking plate to wash out the ethanol from the resin. After centrifugation at 700 rpm for 5 minutes the supernatant was removed, and the resin was again washed with 10 mL of buffer. The lysate was then mixed with the resin and incubated on a shaking table at 4 °C for 50 min, before the mixture was centrifugated at 700 rpm for 5 min and the supernatant (non-binding fraction) removed with a pipette. The resin was then again washed with 10 mL of binding buffer, centrifugated and the supernatant was removed.

The resin was then loaded onto a single-use chromatography column (Poly-Prep® Chromatography Column, Bio-Rad) which retained the resin but allowed flow-through of liquid, unspecific inpuruty were removed by washing the resin with buffers with increasing concentration of imidazole (5mm, 10 mM and 20 mM wash buffer). Finally, his-MsrB1 was eluted with buffer with 250 mM imidazole (elution buffer). In general, the resin was washed with 5 mM imidazole wash buffer until no or little protein could be detected with Bradford Assay in the flow-through (approx. 20-30 mL for 0.5-1.0 mL of resin). A smaller volume of 10 mM and 20 mM imidazole wash buffer, approx. 8 and 2 mL, respectively. Elution buffer was finally added to the column until little protein was detected with Bradford Assay. The

---

[9] Cell lysis and protein purification was performed as described in the Ni-NTA resin manual. Available online at http://www.qiagen.com/literature/render.aspx?id=128. Last visited 13.05.2012.

concentration of the protein in the elution fraction was measured and the solution was stored in a nitrogen (N_2) atmosphere in 4 °C.

3.2.2.3 Purity analysis

To confirm that the protein was induced by IPTG, to check the purity of the protein in the elution fraction and to investigate the efficiency of the different steps of the purification, samples of the all steps of culture growth, expression and purification was analysed by SDS-PAGE gel.

3.2.3 UV-visible spectroscopy

As described in section 1.3.2.2, cobalt(II)-ligand complexes absorb UV-vis light at wavelengths characteristic to the number, nature and coordination geometry of the ligands. UV-vis absorption spectra were therefore recorded for all MsrB1 samples grown in growth media supplemented with $CoCl_2$ to investigate uptake of cobalt by MsrB1.

All protein samples were centrifuged (150,000 rpm, 3 min) prior to recording UV-vis absorption spectra. Approximately 500 μL of protein sample was then transferred to the quartz cuvette and 200-800 nm spectra were always recorded using the same buffer as that of the protein solution as reference. The spectra were baseline-corrected by subtracting the absorption at 800 nm from the absorption at all wavelengths.

3.2.4 pH-dependence of molar extinction coefficient for Co-MsrB1

The characteristic UV-Vis absorption of Co-MsrB1 in the 500-800 nm range is caused by d-d transitions by unpaired d-electrons of the cobalt ion, and changes in this complex such as loss of one ligand should influence the absorption characteristics of Co-MsrB1, best measured by the molar extinction coefficient (ε). It was therefore hypothesized that an investigation of the pH-dependence of ε could reveal the pK_a either of the cobalt-coordinating Cys-residues or nearby pH-titratable residues.

Co-MsrB1 was produced and purified as described in sections 3.2.1.2 and 3.2.2, respectively. Co-MsrB1 in elution buffer was divided into 0.5 mL fractions and dialyzed separately against dialysis buffer 1 with pH values ranging from 5.0 to 12.0. After overnight dialysis, samples of the protein solutions were centrifuged (150 000 rpm, 3 minutes) separately in eppendorf tubes to remove precipitated protein. The pH of the protein solutions were measured to 4.9, 5.4, 5.8, 6.3, 6.8, 7.2, 8.1, 8.2, 8.9, 9.2, 9.4, 10.1, 10.9 and 11.5, respectively. After pH of a sample was measured, UV-vis absorption spectra were immediately recorded and the protein concentration was determined. ε was then calculated for absorption at the three absorption peaks 636, 695 and 723 nm by rearranging the Lambert-Beer law:

$$\varepsilon = I/(bc) \tag{8}$$

where b = 1 cm in these experiments since the diameter of the quartz cuvette is 10 mm, I is the absorption at a given wavelength and c is the protein concentration of the Co-MsrB1 sample. 636, 695 and 723 nm were chosen because these wavelengths constitute the three most distinct peaks of the Co-MsrB1 UV-vis absorption spectrum.

3.2.5 Metal removal from Co- and Zn-MsrB1 and reconstitution

3.2.5.1 Metal removal

To investigate if apo-MsrB1 (metal-free MsrB1) could be produced from Co- and Zn-MsrB1, Co-MsrB1 and Zn-MsrB1 solutions were dialyzed against metal chelating agent containing buffers (EGTA and EDTA) with varying salt and phosphate concentration and different pH-values (5.0, 5.5, 7.0 and 7.5):

2 x 1,6 mL of Co- and Zn-MsrB1 (0.9 and 0.8 mg/mL, respectively) in elution buffer was subjected to extensive dialysis against dialysis buffer 2 with pH = 5.0 and pH = 7.0, respectively, and diluted to 5 mL. The protein solutions were then dialyzed against dialysis buffer 2 containing metal chelating agents EGTA and EDTA (5 mM and 5mM, respectively) overnight for pH = 5.0 and six days for pH = 7.0, at 4 °C stirred at 1 rpm. UV-Vis absorbance spectra were recorded for both Co-MsrB1 samples once per day after overnight dialysis. The Zn-MsrB1 solutions were buffer exchanged into NMR buffer 1 and 1D ^{1}H NMR spectra of the solution were recorded.

8 mL of Co-MsrB1 (3 mg/mL) in elution buffer was dialyzed against dialysis buffer 1 with pH = 5.5 containing EGTA and EDTA (4.2 mM and 5.0 mM, respectively) for eight days at 4 °C stirred with a magnet at 1 rpm. UV-Vis absorbance spectra were recorded of the protein sample once per day after overnight dialysis. To investigate if the dialysis was causing degradation of MsrB1, a sample of the protein solution was analyzed by MALDI-TOF mass spectrometry using SA Double Layer Method (Approximately 0.5 µl of sample and 0.5 µl of sinapinic acid in ethanol were placed twice onto a stainless steel target plate two times and allowed to air-dry before analysis).

2.5 mL of Zn-MsrB1 (2.0 mg/mL) in elution buffer was buffer exchanged into NMR-buffer 1 with pH = 5.5 in a desalting column and mixed with 2.5 mL of NMR-buffer 1 to a total volume of 5 mL and protein concentration of 1 mg/mL. The protein solution was dialyzed against NMR-buffer 1 containing EGTA and EDTA (5 mM and 5 mM, respectively) for five days. 1D ^{1}H NMR spectrum was recorded of the sample after 1, 3 and 5 days.

3.2.5.2 Reconstitution of Co-MsrB1

To attempt reconstitution of Co-MsrB1, apo-MsrB1-solutions (prepared by the procedure described in the previous section) were dialyzed extensively against dialysis buffer 1 or dialysis buffer 2 to remove EDTA and EGTA (three rounds of dialysis, the third overnight). The apo-MsrB1 solutions were then incubated with $CoCl_2$(10 mM $CoCl_2$ added to a 2-6-fold excess ratio to apo-MsrB1) or dialyzed against buffers containing $CoCl_2$ (1-10 µM) in 4 °C in separate experiments for at least 12 hours. Dialysis and incubation was done with pH-values of 5.5 and 7.5 and with and without reducing agent (β-mercaptoethanol) for both buffers. UV-Vis absorbance spectra and 1D ^{1}H NMR spectra were recorded after the treatments to check for metal uptake.

3.2.6 Determination of Co-MsrB1:Zn-MsrB1 ratios in mixed Co- and Zn-MsrB1 samples

In UV-vis absorption spectra, Co-MsrB1 absorbs at characteristic wavelengths while Zn-MsrB1 is spectroscopically silent. The concentration of Co-MsrB1 in a protein sample containing both Zn- and Co-MsrB1 can therefore be found from the UV-vis absorption of these protein samples by rearranging the Lambert-Beer law

$$c = I/(\varepsilon b) \tag{9}$$

when I and ε is known for a given wavelength, since b=1 (1 cm diameter quartz cuvette).

In a protein sample with only Co-MsrB1, c equals the total protein concentration, and ε can then be determined for a given wavelength by rearranging the Lambert-Beer law:

$$\varepsilon = I/(bc) \tag{8}$$

To determine ε for Co-MsrB1, UV-vis absorbance spectra were recorded for three samples of Co-MsrB1 (produced and purified as described in sections 3.2.1.2 and 3.2.2, respectively) in elution buffer and the protein concentrations of the samples were determined. ε was determined for 695 and 723 nm as described in section 3.2.4 using equation (8). 695 and 723 nm were chosen because these wavelengths corresponded to two absorption peaks that gave low variation in the calculated ε.

UV-vis absorption spectra were then recorded for MsrB1 expressed in M9-Co/Zn medium (M9 medium supplemented with 10 µM $CoCl_2$ and 10 µM $ZnSO_4$, four samples) and MsrB1 expressed in LB medium supplemented with 10 µM $CoCl_2$ and 50 µM $CoCl_2$ (protein produced as described in sections 3.2.1.2 and 3.2.1.3, respectively and purified as described in section 3.2.2) in elution buffer, and the protein concentration of the three protein solutions was determined. For each sample, the concentration of Co-MsrB1 (c) of the sample was determined by equation (9) using the absorptions for these samples at 695 and 723 nm. The Co-MsrB1:Zn-MsrB1 concentration ratio was found by dividing the calculated concentration of Co-MsrB1 (c) in the sample with the total protein concentration sample (c_{tot}):

$$\text{Co-MsrB1:Zn-MsrB1} = c/c_{tot} \tag{10}$$

The ratio was calculated for each ε (from three Co-MsrB1 samples, two wavelengths), thus giving six calculated ratios for each Zn- and Co-MsrB1 sample. The average of the calculated ratios was found, and the standard deviation for the ratio found using the STDEV.P-function in Microsoft Excel 2010.

The method is summarized in Figure 3.1.
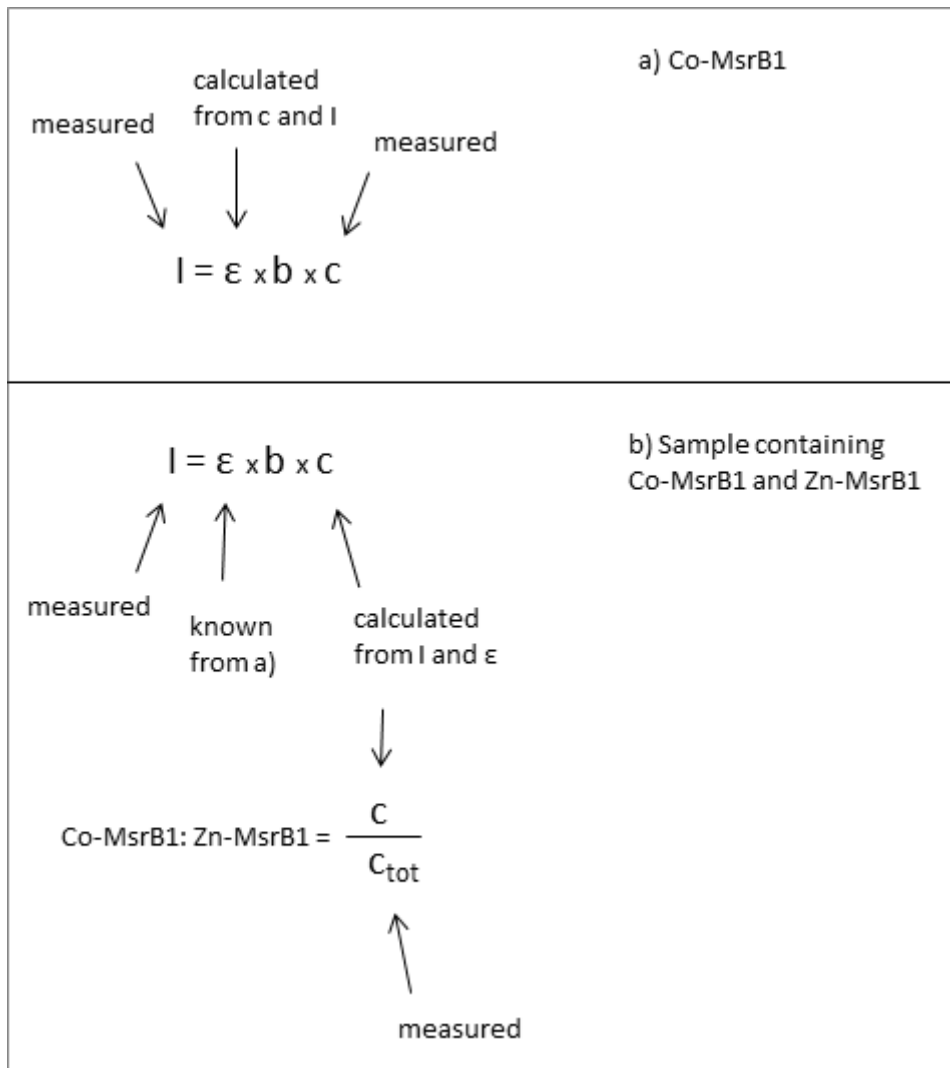
Figure 3.1: Summary of Co-MsrB1:Zn-MsrB1 ratio determination method. a) For samples containing only Co-MsrB1, the absorption (I) at a given wavelength and concentration of Co-MsrB1 (c) can be measured, allowing calculation of ε. b) For samples of MsrB1 expressed in zinc- and cobalt-supplemented growth media, both Zn- and Co-MsrB1 is present in the purified protein, and only I can be measured for Co-MsrB1 of the sample. Using ε calculated from a), c can be determined. b = 1 for all calculations, as the diameter of the quartz-cuvette is 1 cm.

Section 3.2.1 to 3.2.6, constituting the biological and biochemical part of this study, is summarized as a flowchart in Figure 3.2.
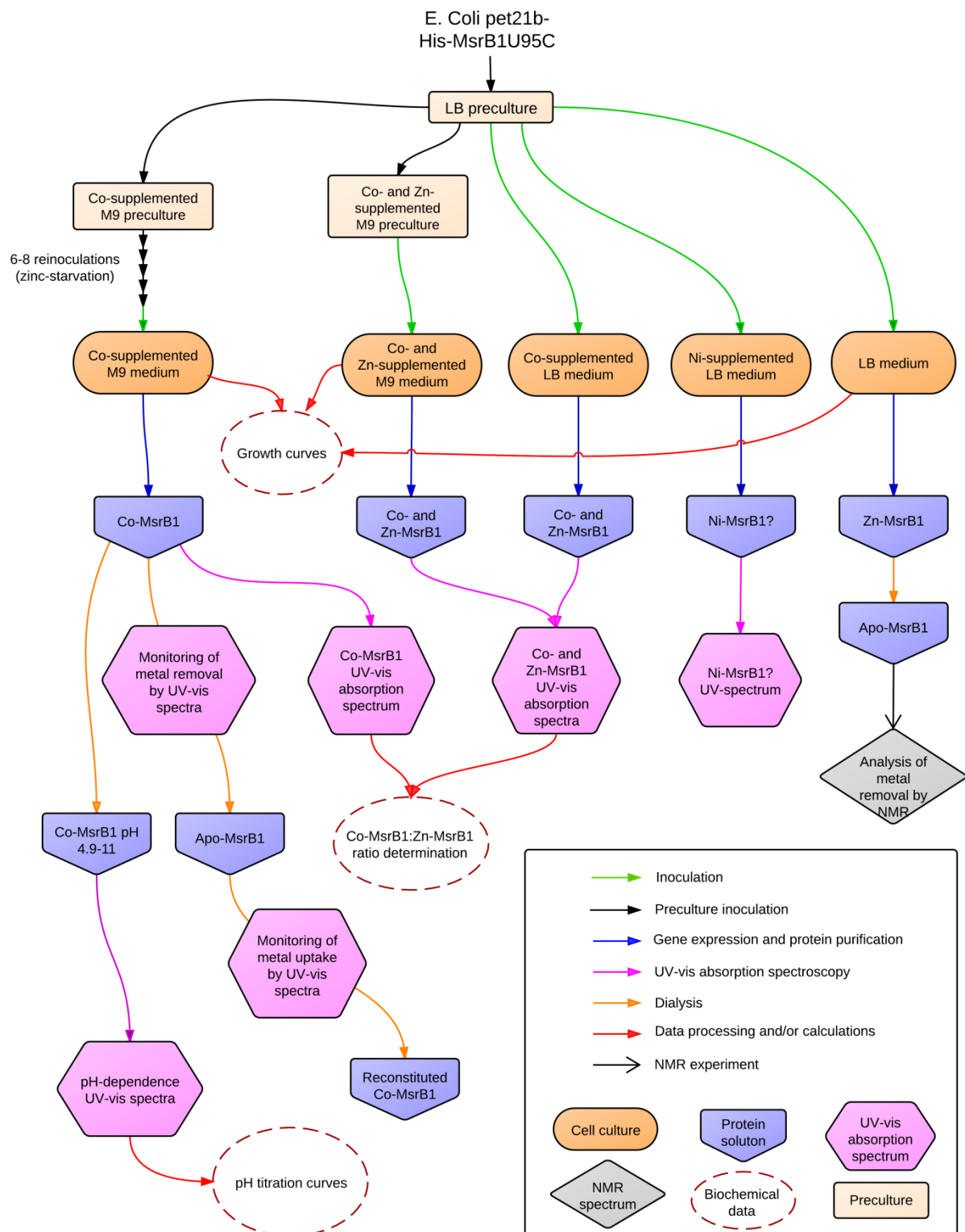
Figure 3.2: Flowchart representation of sections 3.2.1-3.2.6. Experimental procedures of sections 3.2.1-3.2.6 summarized as a flowchart. All experiments are initiated by inoculation of an LB-preculture from the frozen 15 % glycerol stock-solution of *E. Coli* carrying a pET-21b His-MsrB1u95C construct (top). Chart interpretation-key is shown in a separate box (bottom right).

3.2.7 NMR spectroscopy

3.2.7.1 Preparation of protein sample for NMR experiments

Before NMR experiments were performed, protein samples were buffer exchanged into NMR buffer 1 or 2 by dialysis or a centrifugal filter (Amicon® Ultra, 10,000 MWCO, Millipore) that retained the protein but allowed flow-through of buffer. 450 µL of the sample was then transferred to an eppendorf tube and 50 µL D_2O (99.9%, CDN Isotopes) was added. The sample was then mixed gently and transferred to an NMR tube.

3.2.7.2 NMR-experiments

Below, the different NMR spectra used in this thesis and their origin are described.

The following NMR-spectra of Co-MsrB1 were recorded at 298 K on an Avance 800-MHz equipped with a 5-mm z-gradient CP-TCI(H/C/N) cryoprobe at the Magnetic Resonance Centre (CERM) in Florence, Italy:

- 1D ^1H (large spectral windows), ^{15}N HSQC (large and small spectral windows), CBCA(CO)NH (large and small spectral windows), CBCANH (large and small spectral windows), HAHB(CO)NH (large and small spectral windows), HAHBNH (large and small spectral windows), NH(CA)CO (large and small spectral windows), NHCA (large and small spectral windows), NHCO (large and small spectral windows)

The following NMR-spectra of Zn-MsrB1 were recorded at 298 K on an Avance 800-MHz equipped with a 5-mm z-gradient CP-TCI(H/C/N) cryoprobe at the Magnetic Resonance Centre (CERM) in Florence, Italy:

- 1D ^1H (large spectral windows)

The following NMR spectra of Co-MsrB1 were recorded at 298 K on a Bruker Avance 600 MHz spectrometer equipped with a 5-mm z-gradient CP-TCI(H/C/N) cryoprobe at the NT faculty NMR center at NTNU, Gløshaugen:

- 1D ^1H (small spectral window), ^{15}N HSQC (D_2O), HCCH-TOCSY

The spectra were acquired using pulse sequence from the standard pulse sequence library. The NMR data were processed with the BRUKER TopSpin version 2.1.

3.2.7.3 NMR assignment

^{1}H, ^{15}N, ^{13}C backbone resonance assignments for MsrB1-Co were achieved with the ^{15}N-HSQC, ^{13}C-HSQC, HNCA, HNCO, CACB(CO)NH, CACBNH, HBHANH, HBHA(CO)NH NMR spectra by Computer Aided Resonance Assignment (CARA) based on the CARA Tutorial[10]. The backbone-assignment was initiated by identification and assignment of spin systems to the cross peaks of the ^{15}N HSQC spectrum in the CARA application PolyScope, and the 3D triple resonance spectra were used for further assignment of backbone- and side chain-atoms and to find and link the spin systems corresponding to amino acids adjacent in the primary sequence. To identify which amino acids a stretch of linked spin systems belonged to, a database[11] of typical chemical shifts for the atoms of amino acids was consulted. Pro-residues were identified by characteristic C^{α}_{i-1} and C^{β}_{i-1} chemical shifts for the assigned spin system.

Since a full assignment of ^{1}H, ^{13}C and ^{15}N atoms has been completed for Zn-MsrB1 (BioMagResBank accession no. 15193 or [44]), these data were continuously consulted and compared. As the assignment of Co-MsrB1 progressed, the pseudocontact shifts were calculated for backbone H^N-, N- and C^{α}- and C^{β}-atoms based on the chemical shifts of these atoms for Zn-MsrB1 by equation (4):

$$\delta^{PCS} = \delta^{OBS} - \delta^{dia} \qquad (4)$$

The structure of Zn-MsrB1 has also been solved with NMR (PDB ID 2KV1). Working with a hypothesis that Co simply replaces Zn in Co-MsrB1 and thus leads to little structural change in the protein, the interatomic distances from Zn to backbone H^N- and N-atoms were calculated for conformer 1 of Zn-MsrB1 and used as theoretical distances from Co to the same atoms. A plot of pseudo contact shift vs. theoretical distance was used to evaluate the working hypothesis, identify assignments that were likely erroneous and to predict chemical shifts for backbone N- and HN-atoms of unassigned residues.

^{1}H, ^{13}C side chain assignment was achieved using the HCCH-TOCSY NMR spectrum by the CARA application Systemscope based on the backbone assignment. This assignment was also used to confirm the backbone assignment.

The ^{15}N HSQC spectrum of Co-MsrB1 in D_2O was overlayed with the ^{15}N HSQC assignment of Co-MsrB1 and residues to which signals of Co-MsrB1 in D_2O belonged were identified.

---

[10] Tutorial by Rochus L. J. Keller, Institute for Molecular Biology and Biophysics, The Swiss Federal Institute of Technology, Zürich, Switzerland, 2004, 1st ed. Available online at http://cara.nmr-software.org/downloads/3-85600-112-3.pdf. Further material on CARA is available at http://cara.nmr.ch/doku.php.

[11] Statistics Calculated for All Chemical Shifts from Atoms in the 20 Common Amino Acids, Biological Magnetic Resonance Data Bank. Available online at http://www.bmrb.wisc.edu/ref_info/statful.htm. Last visited 18.04.12.

The large spectral window 1D ^{1}H spectra of Zn-MsrB1 and Co-MsrB1 were compared for identification of paramagnetically shifted signals of Co-MsrB1.

3.2.7.4 Elaboration of NMR data

As described in section 1.3.1.3, PCSs can be used to create a mathematical model of the magnetic susceptibility which allows for evaluation of a (preliminary) protein structure by comparing observed PCSs with PCSs calculated with the model. Since a structure of Zn-MsrB1 is available, the method could be used to test both the quality of the NMR assignment and investigate if cobalt simply takes the place of zinc in Co-MsrB1. From the pseudocontact shifts (PCSs) for H^N of the assigned residues, a magnetic susceptibility-tensor for conformer 1, 2, 3 and the mean (average) conformer of the published structure for MsrB1 was calculated by Rebecca del Conte (CERM) using the software AnisoFit[12] (previously known as Fantasia [23]). From the calculated tensors, PCSs were calculated for the assigned residues and predicted for unassigned residues within a 7 Å and between 7-10 Å distance, respectively, from zinc in the published structure.

As already mentioned in the previous section, plots of PCSs for H^N vs. hypothetical distance to cobalt were made for the assigned residues using conformer 1 in order to evaluate single residue HSQC assignments and to visualize the reverse proportional relationship between H^N PCS and distance to cobalt (PCS $\propto$ 1/r^3). To further investigate the PCS-distance-relationship in order to evaluate the structure of Co-MsrB1, interatomic distances from the Zn ion to all backbone H^N was calculated for the remaining 19 conformers of the published structure for MsrB1. For each conformer, a plot was made in the same way as for conformer 1. PCS vs. distance plots for all conformers were also overlaid to see if a clear distance-PCS relationship appeared.

To investigate what changes would occur in these plots if the position of the cobalt metal ion was slightly shifted, representing a small structural change within the protein, the interatomic distances of C' of K25 (which resides between two metal-coordinating Cys-residues in Zn-MsrB1) were calculated, and a plot was made of PCSs for H^N vs. the distance to this atom. In addition, interatomic distances from all H^N to the first His-residue of the His-tag (LE*H*HHHHH) were calculated, and a plot was made of the PCSs for H^N vs. the respective H^N-His distances. This was done to investigate how a PCS vs. distance plot would look for what was assumed to be the wrong distances to Co(II)-ion, in other words to create a "non-sense" plot, and to show that the PCS-data supported that the His-tag is not responsible for uptake of cobalt.

---

[12] AnisoFit is available online at http://alchemy.cchem.berkeley.edu/anisofit2.0/anisofit.html.

Sections 3.2.7.2-3.2.7.4 are summarized as a flowchart in Figure 3.3.
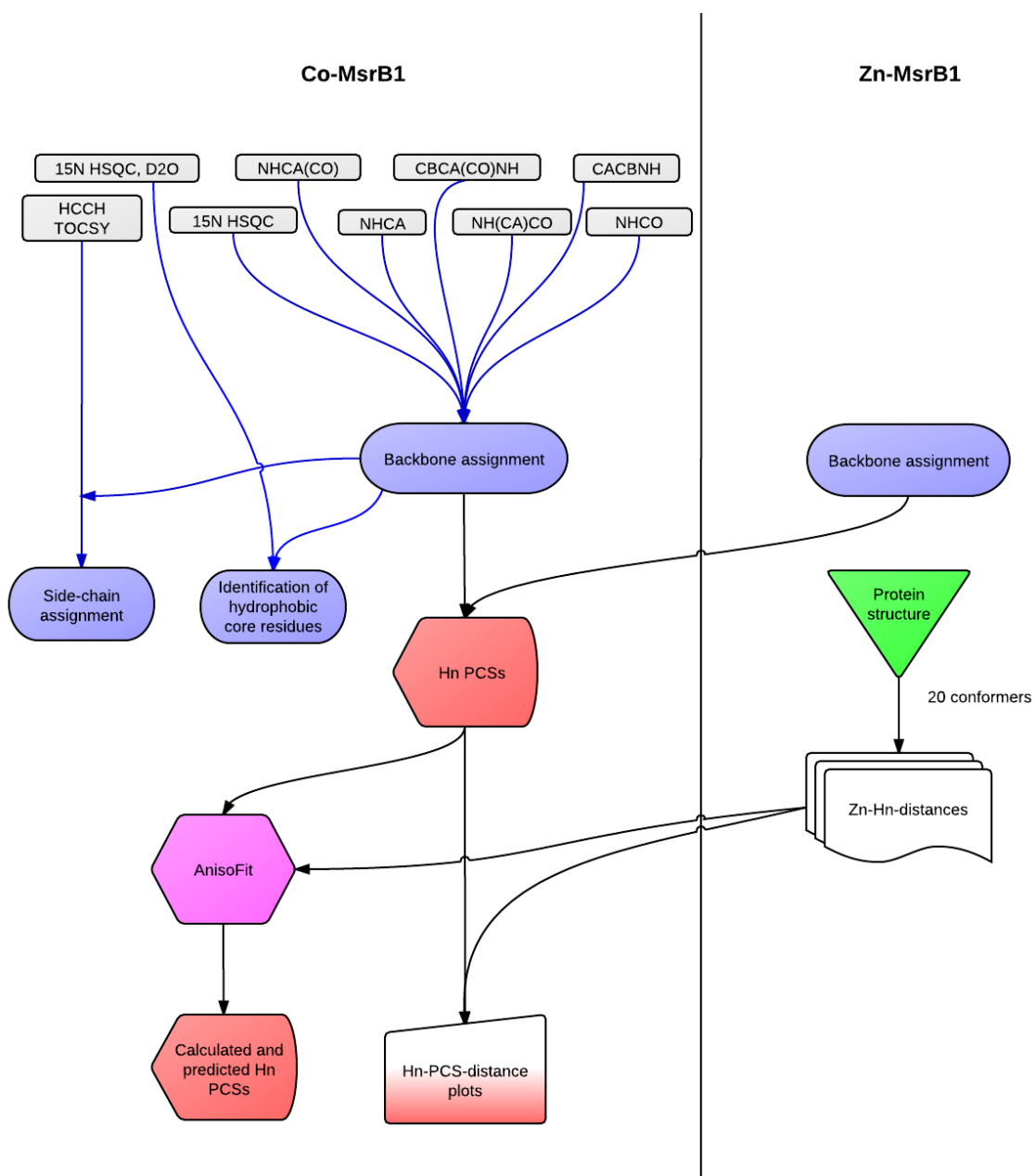
Figure 3.3: Flowchart representation of sections 3.2.7.2-3.2.7.4. Experimental procedures of sections 3.2.7.2-3.2.7.4 (except 1D experiments) summarized as a flowchart. Blue arrows represent assignment of signals and black arrows represent calculations and/or data processing.

4 Results

4.1 Effects of cobalt(II) on culture growth

Cultures of *E. Coli* were grown in LB medium, zinc-supplemented M9 medium with different concentrations of cobalt and in cobalt-supplemented M9 medium after zinc-starvation of the bacteria, as described in sections 3.2.1.1, 3.2.1.2 and 3.2.1.5, respectively. Growth in LB-medium gave the highest growth rate. For zinc- and cobalt-substituted M9 media growth rates decreased with increasing concentration of $CoCl_2$, and for concentrations of 50 µM or above a leveling off from logarithmic to linear growth was observed. Cultures in cobalt-supplemented M9 medium with zinc-starved *E. Coli* showed logarithmic growth with a later onset than for media containing Zn. The recorded growth curves for the different media are shown in Figure 4.1.
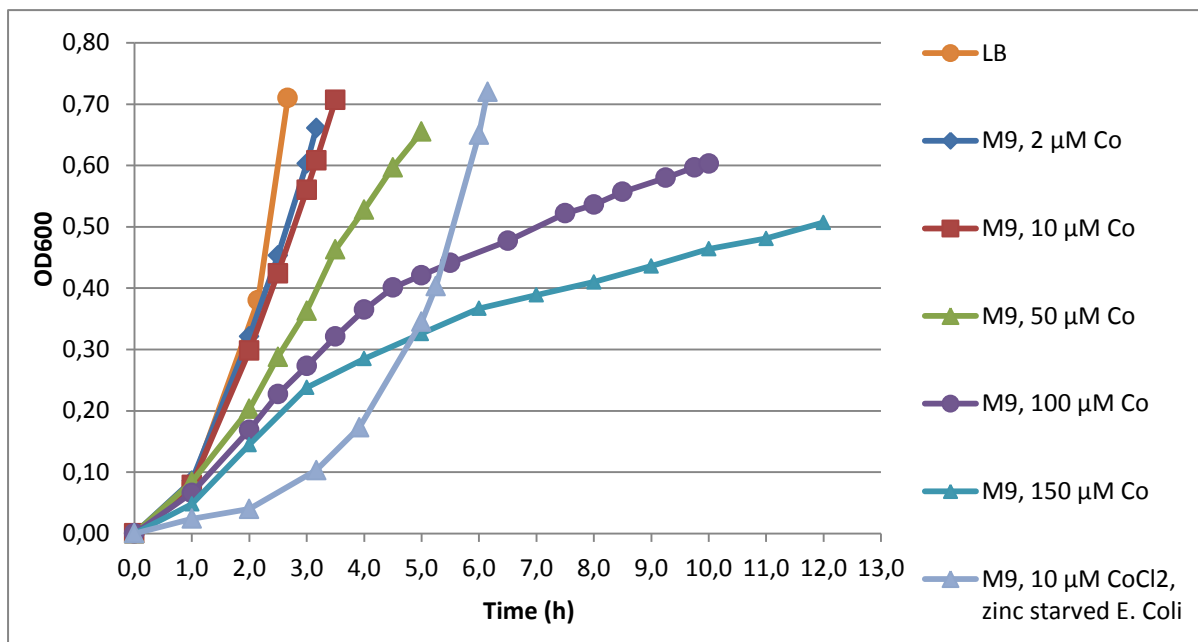
Figure 4.1: Growth curves for different growth media: Plot of OD_{600} as a function of time for *E. Coli* in LB, M9-media with 10 µM $ZnSO_4$ and 2-150 µM $CoCl_2$ and Zn-depleted *E.*Coli in M9-Co medium. The diagram shows that the growth rates decrease with increasing concentration of $CoCl_2$, and that Zn-depleted *E. Coli* in Zn-free medium has logarithmic growth but with a later onset.

4.2 Protein expression and purification of MsrB1

Protein production and purification of MsrB1 was carried out as described in section 3.2.1, and samples from the different steps from culture growth to eluted protein were analyzed by SDS-PAGE as described in section 3.2.2.3. As seen in Figure 4.2, elution fractions had a strong band at about 14 kda, corresponding to the molecular weight of MsrB1 (see Appendix A).
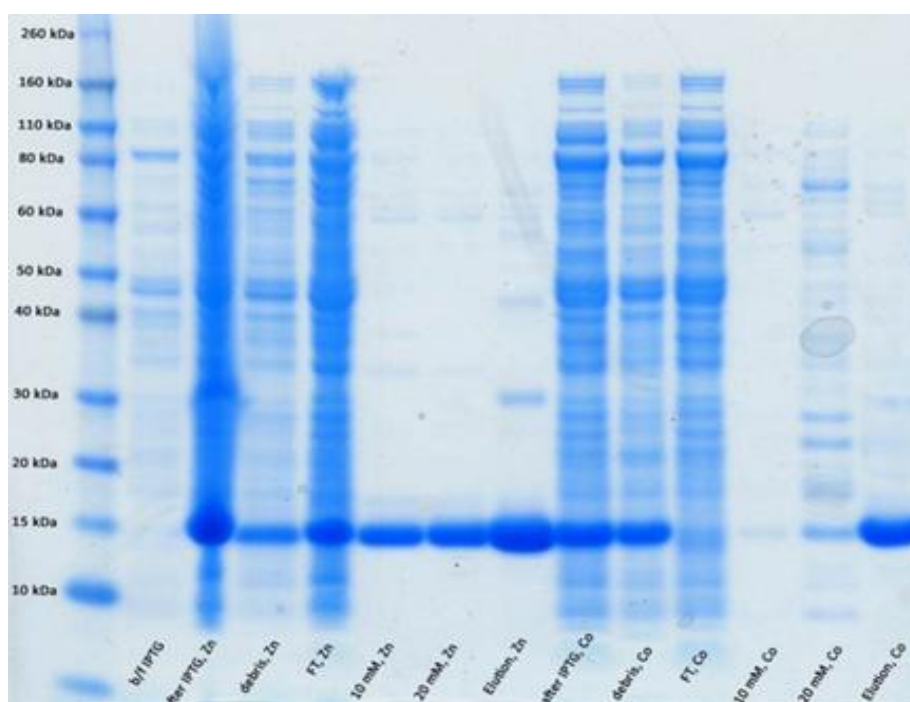
Figure 4.2: SDS-PAGE: Samples of the different steps from culture growth to purified MsrB1 analyzed by SDS-PAGE, with labels indicating fragment size of protein standard and content of the individual wells. *b/f IPTG* and *after IPTG* refers to cell extract from growth culture before and after induction of gene expression with IPTG, respectively. *Debris* refers to the insoluble fraction of lysed cells. *FT, 10 mm, 20 mm* and *elution* refers to fractions of 5 mM, 10 mM, 20 mM and 250 mM of imidazole buffers collected from the Ni-NTA column, respectively. *Zn* and *Co* indicates Zn- and Co-MsrB1, respectively. The well to the far right is non-related. The bands around 15 kDa indicate the presence of MsrB1 in all fractions after, but not before, induction of gene expression. In the elution fractions of both Zn- and Co-MsrB1, slight impurities are visible as bands at approx. 30 and 45 kDa.

The production of MsrB1 in 2 L of bacterial culture typically gave 8-10 mL of elution fraction with a protein concentration of 2-3 mg/mL, irrespective of growth medium, thus giving a protein yield of 8-15 mg/L of bacterial culture. Elution fractions of Zn-MsrB1 were colorless or slightly yellow while elution fractions of Co-MsrB1 had a strong blue color. Both forms of MsrB1 were stable and could be stored, even at higher concentrations, without precipitating.

4.3 Metal uptake by MsrB1

Cultures of *E. Coli* were grown as described in section 3.2.1 with Co- and Ni-supplemented media as described in sections 3.2.1.2-3.2.1.4, and expression and purification of MsrB1 was performed as described in sections 3.2.1 and 3.2.2, respectively. UV-vis absorption spectra of the purified protein solutions were recorded as described in section 3.2.3.

Uptake of cobalt(II) by MsrB1 was detected with UV-vis spectroscopy for all cobalt(II)-supplemented growth media, indicated by absorption peaks observed at about 335, 631, 660, 695 and 723 nm owing to coordinated cobalt(II). Uptake of nickel by MsrB1 was not detected (not shown). UV-vis absorption spectra for the purified protein from different media are shown in Figures 4.3 and 4.4.
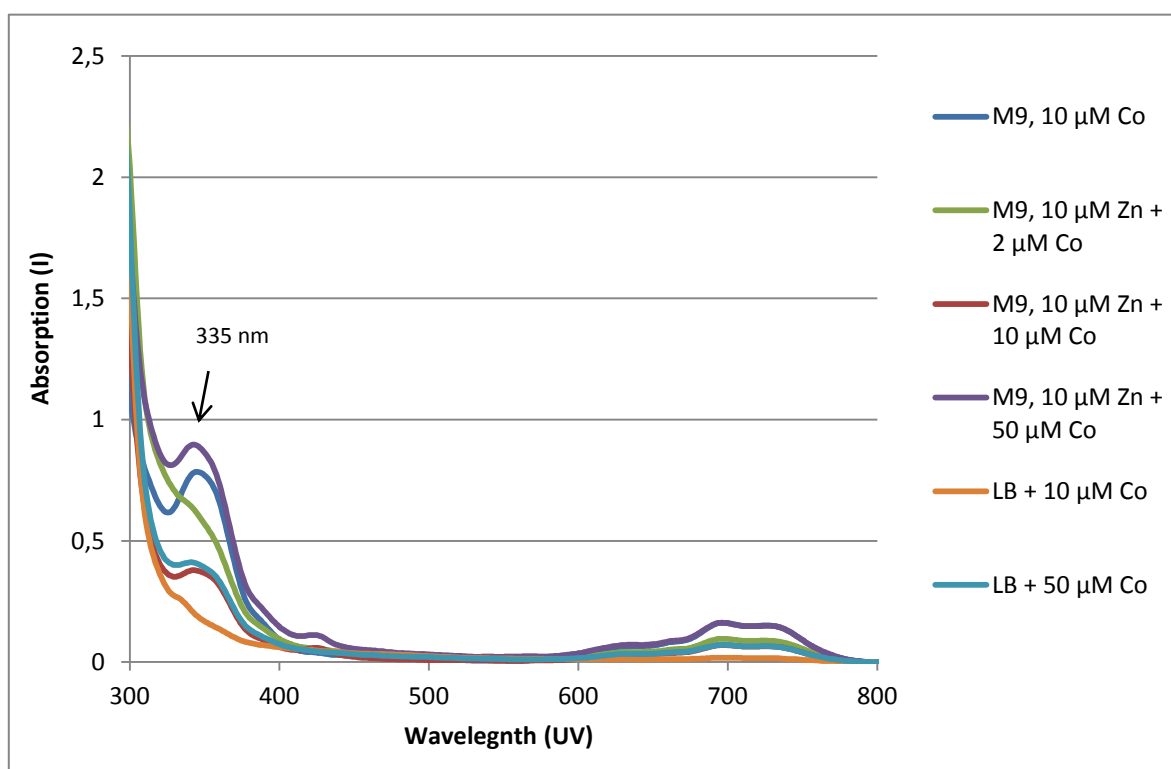
Figure 4.3: UV-vis absorption spectra of Co-MsrB1 produced with different growth media: Absorption (I) shown as a function of wavelength (nm) for MsrB1 purified from cell cultures grown in different media supplemented with $CoCl_2$. The characteristic absorption profiles from at approx. 340 and 600-800 nm show cobalt-uptake by MsrB1.
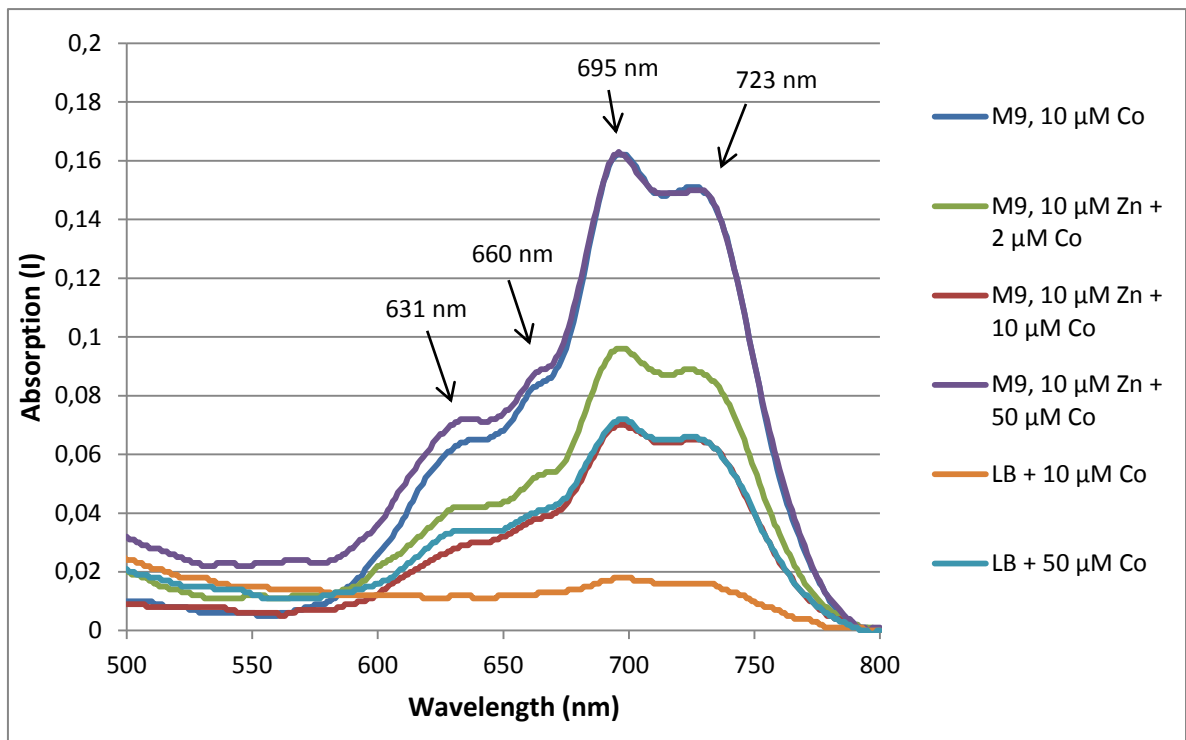
Figure 4.4: UV-vis absorption spectra (500-800 nm) of Co-MsrB1 produced with different growth media: Absorption (I) shown as a function of wavelength (nm) for MsrB1 purified from cell cultures grown in different media supplemented with $CoCl_2$. The characteristic absorption profile shows uptake of cobalt by MsrB1. Characteristic absorption peaks and are indicated. Note that the solutions have different protein concentration, and the band intensities therefore only reflect the Co-MsrB1 concentrations of the respective samples.

4.4 pH-dependence of the molar extinction coefficient for Co-MsrB1

pH-adjustment of Co-MsrB1 and calculation of the extinction coefficient (ε) was carried out as described in section 3.2.4. The calculated ε together with data used in calculations are shown in table 4.1. UV-vis absorption spectra of Co-MsrB1 in pH 4.9-11.5 are shown in Appendix C.

Table 4.1: ε calculated for Co-MsrB1 at different pH. Absorption at the three wavelengths used as probe are shown for the different pH, together with the protein concentration of the sample and the calculated ε.

pH	Absorption, I			c (mg/mL)	ε		
	636 nm	695 nm	723 nm		636 nm	695 nm	723 nm
11,54	0,009	0,010	0,008	1,44	0,00625	0,00694	0,00556
10,87	0,024	0,058	0,056	1,17	0,02051	0,04957	0,04786
10,1	0,028	0,065	0,060	1,21	0,02314	0,05372	0,04959
9,37	0,033	0,073	0,068	1,28	0,02578	0,05703	0,05313
9,21	0,027	0,063	0,059	1,13	0,02389	0,05575	0,05221
8,87	0,027	0,060	0,055	1,08	0,02500	0,05556	0,05093
8,19	0,028	0,065	0,061	1,21	0,02314	0,05372	0,05041
7,19	0,018	0,043	0,040	0,99	0,01818	0,04343	0,04040
6,78	0,029	0,070	0,066	0,86	0,03372	0,08140	0,07674
6,32	0,034	0,081	0,077	1,00	0,03400	0,08100	0,07700
5,84	0,025	0,057	0,054	0,80	0,03125	0,07130	0,06750
5,42	0,015	0,028	0,025	1,09	0,01376	0,02569	0,02294
4,92	0,007	0,002	0,001	1,18	0,00593	0,00170	0,00085

ε showed maxima at pH 6.8 and 9.4 or 6.3 and 9.4, depending on the wavelength used as probe, and rapidly decreases to zero for pH lower than 6.3 and higher than 10.9. The low pH maximum is part of a narrow peak of ε, while the high pH maximum is bell shaped (Figure 4.5). The wavelengths of the absorption peaks and the overall absorption profile remained constant for all pH-values.
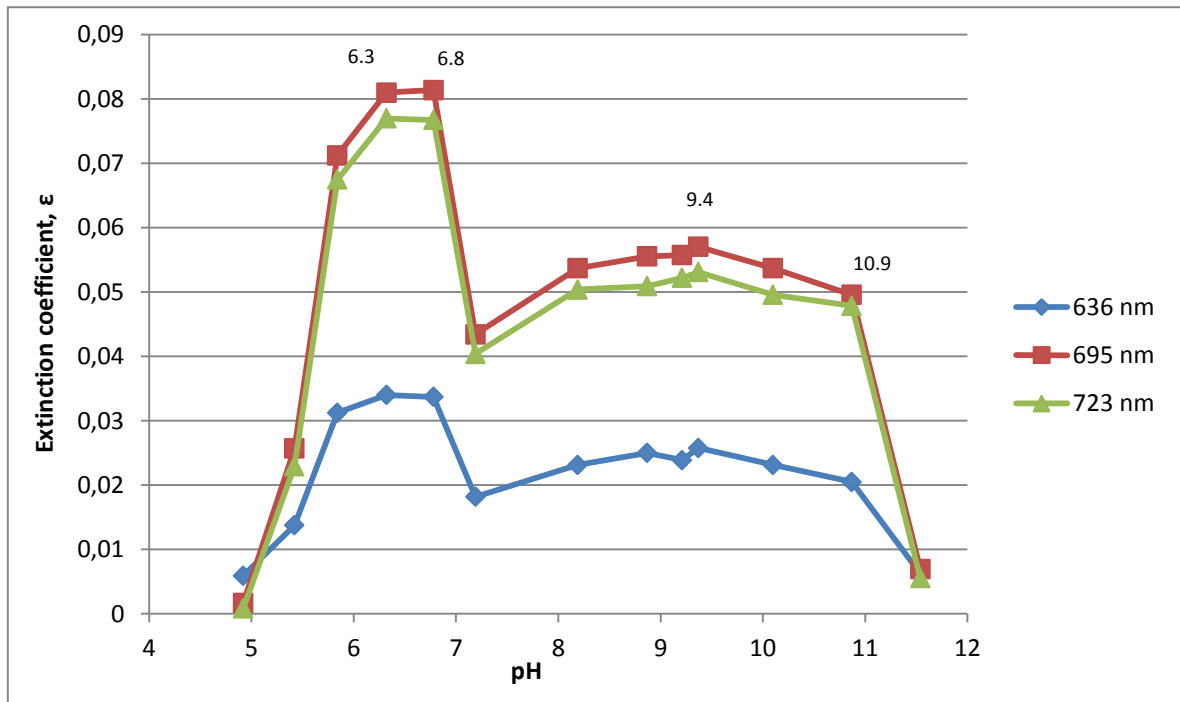
Figure 4.5: pH dependence of εb for Co-MsrB1: Plot of the extinction coefficient (ε) of Co-MsrB1 as a function of pH, with pH-values labeled for data points mentioned in the text.

4.5 Metal removal from Co- and Zn-MsrB1 and reconstitution

Samples of Zn-MsrB1 and Co-MsrB1 were dialyzed against chelating agents as described in section 3.2.5.1. Apo-MsrB1 was dialyzed against and incubated with buffers containing $CoCl_2$ to reconstitute Co-MsrB1 as described in section 3.2.5.2.

Co was successfully removed from Co-MsrB1 at pH 5.0 and 5.5 (see Figures 4.6, 4.7 and 4.8), but unsuccessful at pH 7.0. Analysis of the protein solutions by MALDI-TOF mass spectrometry showed that the protein was not degraded in the process (not shown). Removal of Zn from Zn-MsrB1 was not achieved, as analyzed with 1D ^{1}H NMR spectroscopy (results not shown). Reconstitution of Co-MsrB1 was not detected on any occasion with either UV-vis spectroscopy or 1D ^{1}H NMR spectroscopy (not shown).
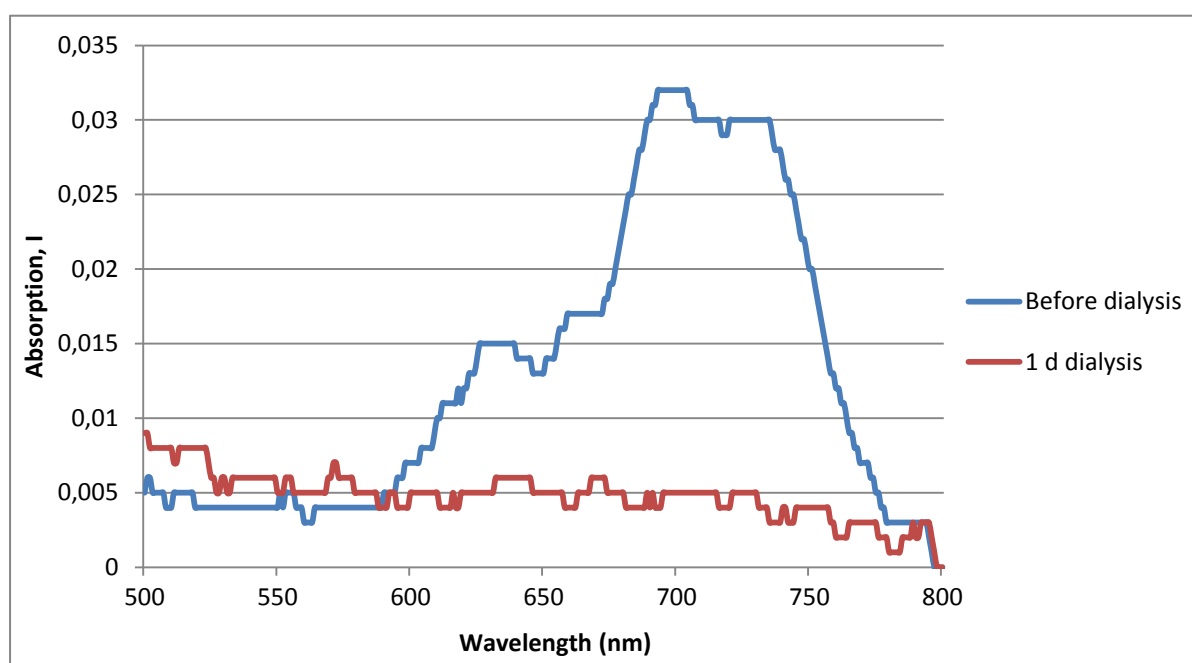
Figure 4.6: UV-vis spectrum (500-800 nm) of Co-MsrB1 and apo-MsrB1: UV-vis absorption (I) as a function of wavelength (nm) in the 500-800 nm range for Co-MsrB1 before and after overnight dialysis against chelating agents EGTA and EDTA (5mM and 5mM, respectively). The absorption peaks characteristic for bound Co have completely disappeared, indicating that Co has been removed from the protein.
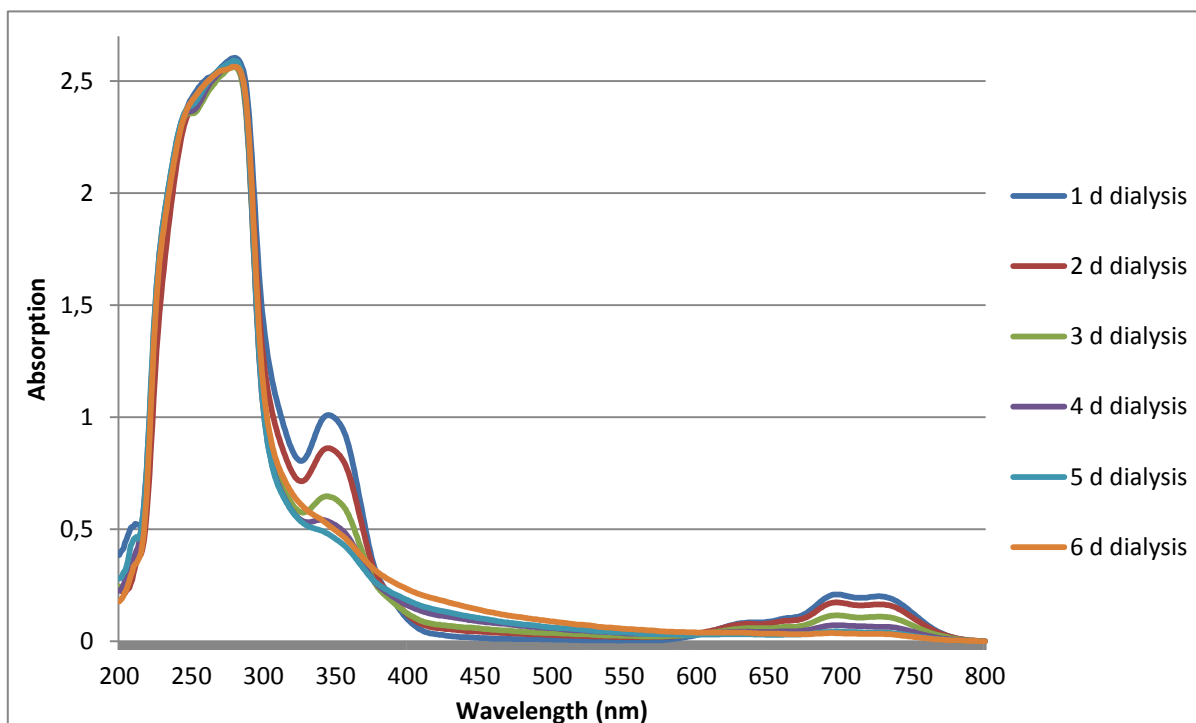
Figure 4.7: UV-vis spectrum (200-800 nm) of Co-MsrB1 during metal removal: UV-vis absorption (I) as a function of wavelength (nm) in the 200-800 nm range for Co-MsrB1 after overnight and five following days of dialysis against chelating agents EGTA and EDTA (4mM and 5mM, respectively). The spectra show that the absorption peaks characteristic for bound Co slowly diminish, indicating gradual loss of Co for Co-MsrB1.
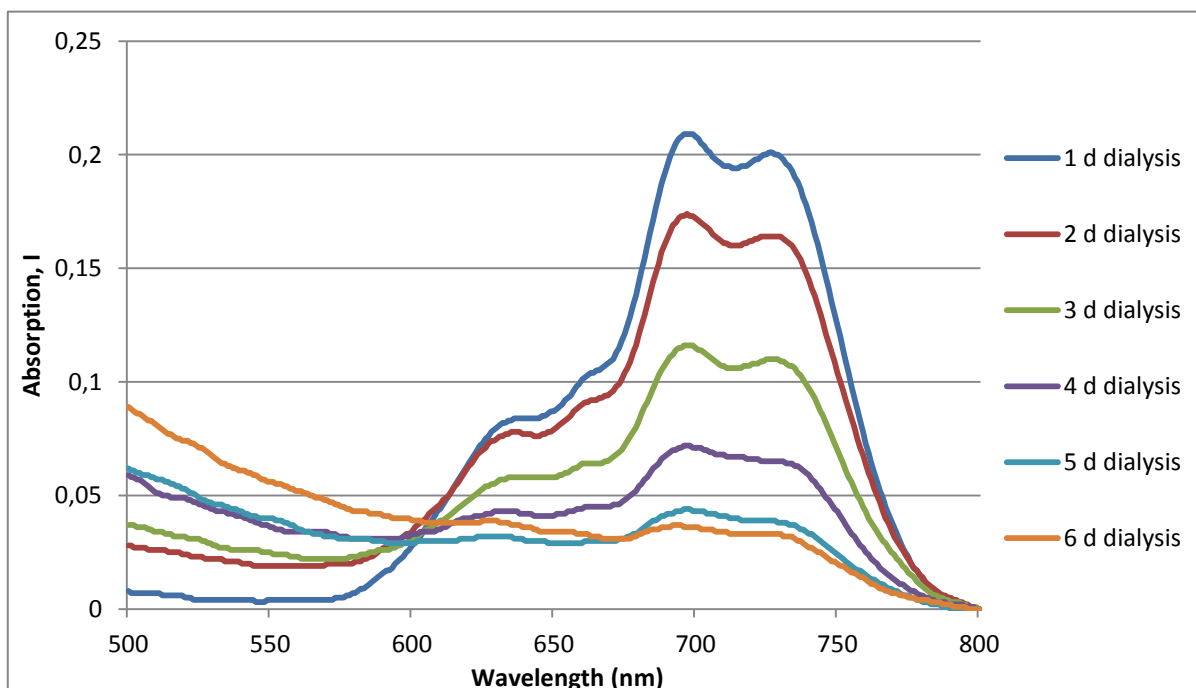
Fig. 4.8: UV-vis spectrum (500-800 nm) of Co-MsrB1 during metal removal: Enlargement of the 500-800 nm area of the spectra shown in Fig. 4.7.

4.6 Determination of Co-MsrB1:Zn-MsrB1 ratios in mixed Co- and Zn-MsrB1 samples

The molar extinction coefficient (ε) for Co-MsrB1 and Co-MsrB1:Zn-MsrB1 ratios for MsrB1 expressed in cobalt- and zinc-supplemented growth media were calculated as described in section 3.2.6. The Co-MsrB1:Zn-MsrB1 ratios in M9 medium with 10 µM $ZnSO_4$ and 10 µM CoCl2, LB supplemented 50 µM $CoCl_2$ and LB supplemented with 10 µM $CoCl_2$ were found to be 0.184, 0.111 and 0.032, respectively. Calculations of ε for Co-MsrB1 and Co-MsrB1:Zn-MsrB1 ratios are shown in Tables 4.1 and 4.2, respectively. The Co-MsrB1:Zn-MsrB1 ratios are graphically displayed in Figure 4.9.

Table 4.1: Calculations of ε for Co-MsrB1. Absorption, concentration and calculated ε for Co-MsrB1.

Sample	Concentration of Co-MsrB1 (mg/mL)	Abs. (I), 695 nm	Abs. (I), 723 nm	ε	Calculated molar extinction coefficient (ε)	
					695 nm	723 nm
Co-MsrB1 1	10,3	0,878	0,82	ε 1	0,0852	0,0796
Co-MsrB1 2	2,07	0,162	0,151	ε 2	0,0783	0,0729
Co-MsrB1 3	1,77	0,143	0,134	ε 3	0,0808	0,0757

Table 4.2: Calculations of Co:MsrB1-Zn-MsrB1 ratios. Co-MsrB1 concentration and ratio of Co-MsrB1:Zn-MsrB1 calculated for mixed protein samples using ε from table 4.1, absorption at 695 and 723 nm and the total protein concentration of the samples.

Sample	Total sample protein conc. (mg/mL)	Abs. (I), 695 nm	Abs. (I), 723 nm	ε	Concentration of Co-MsrB1 in sample (mg/mL) 695 nm	723 nm	Ratio of Co-MsrB1:Zn-MsrB1 in sample (mg/mL) 695 nm	723 nm	Ratio average	Std. dev.
M9 Co/Zn 1	2,02	0,032	0,028	ε 1	0,3754	0,3517	0,1858	0,1741		
				ε 2	0,4089	0,3838	0,2024	0,1900		
				ε 3	0,3961	0,3699	0,1961	0,1831		
M9 Co/Zn 2	2,005	0,031	0,028	ε 1	0,3637	0,3517	0,1814	0,1754		
				ε 2	0,3961	0,3838	0,1976	0,1914		
				ε 3	0,3837	0,3699	0,1914	0,1845	**0,1839**	**0,0084**
M9 Co/Zn 3	4,8	0,07	0,065	ε 1	0,8212	0,8165	0,1711	0,1701		
				ε 2	0,8944	0,8911	0,1863	0,1856		
				ε 3	0,8664	0,8586	0,1805	0,1789		
M9 Co/Zn 4	4,73	0,07	0,065	ε 1	0,8212	0,8165	0,1736	0,1726		
				ε 2	0,8944	0,8911	0,1891	0,1884		
				ε 3	0,8664	0,8586	0,1832	0,1815		
LB 50 µM Co	7,87	0,071	0,066	ε 1	0,8329	0,8290	0,1058	0,1053		
				ε 2	0,9072	0,9048	0,1153	0,1150	**0,1106**	**0,0039**
				ε 3	0,8788	0,8718	0,1117	0,1108		
LB 10 µM Co	6,85	0,018	0,016	ε 1	0,2112	0,2010	0,0308	0,0293		
				ε 2	0,2300	0,2193	0,0336	0,0320	**0,0315**	**0,0014**
				ε 3	0,2228	0,2113	0,0325	0,0309		
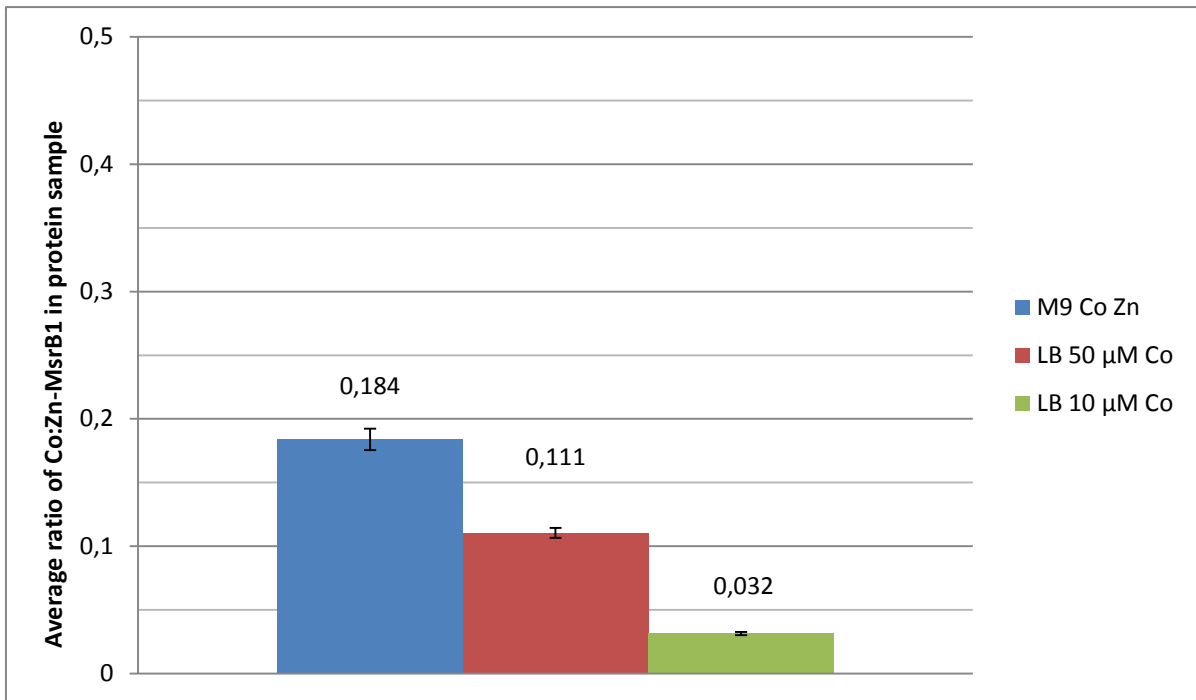
Figure 4.9: Co-MsrB1:Zn-MsrB1 ratios in different growth medium. The Co-MsrB1:Zn-MsrB1 ratios of purified MsrB1 expressed in M9-medium with 10 μM CoCl$_2$, LB-medium supplemented with 50 μM CoCl$_2$ and LB-medium supplemented with 10 μM CoCl$_2$. The determined ratios are represented as bars with a label above each bar representing the determined ratios, and the standard deviation is shown as black error bars.

4.7 Co-MsrB1 NMR spectra and assignment

4.7.1 NMR spectra

NMR spectra of Co-MsrB1 were obtained and processed as described in section 3.2.7.2.

The ^{15}N HSQC spectrum showed discreet peaks, showing that Co-MsrB1 is a folded protein with stable structure (see Figures 4.10 and 4.11) and comparison with the ^{15}N HSQC for Zn-MsrB1 showed that the two spectra are unique (see Figure 4.12).

^{15}N HSQC spectrum for Co-MsrB1 D$_2$O showed that some residues have slow backbone H^{N}-deuterium exchange (Figure 4.13).

Comparison of small and large spectral windows 1D ^{1}H spectra of Zn-MsrB1 and Co-MsrB1 revealed several signals shifted above 15 ppm and below 4 ppm for Co-MsrB1 (outside of the processed ^{15}HSQC spectrum), some signals as far as 350 ppm (see Figures 4.14-4.17).
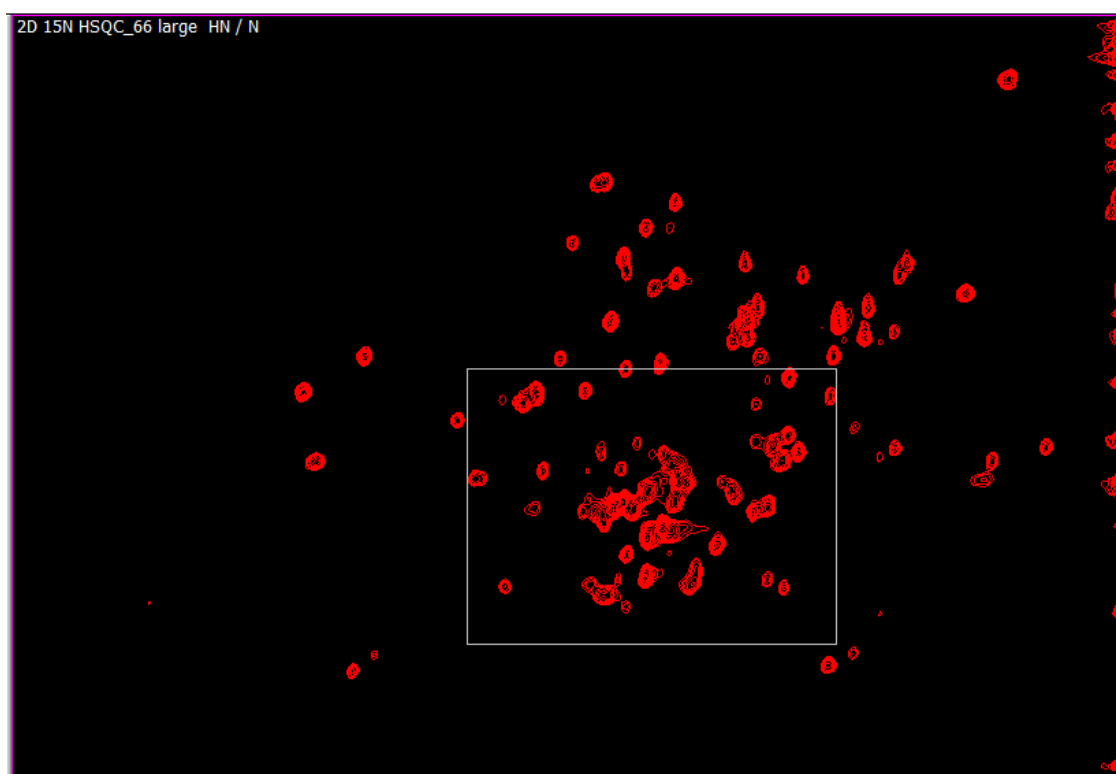
Figure 4.10: ^{15}N HSQC of Co-MsrB1: Section of the ^{15}N HSQC spectrum of Co-MsrB1 where cross peaks were observed. The discreet peaks demonstrate that Co-MsrB1 is a folded protein with stable structure. A white square in the centre indicates the area enlarged in Figure 4.11.
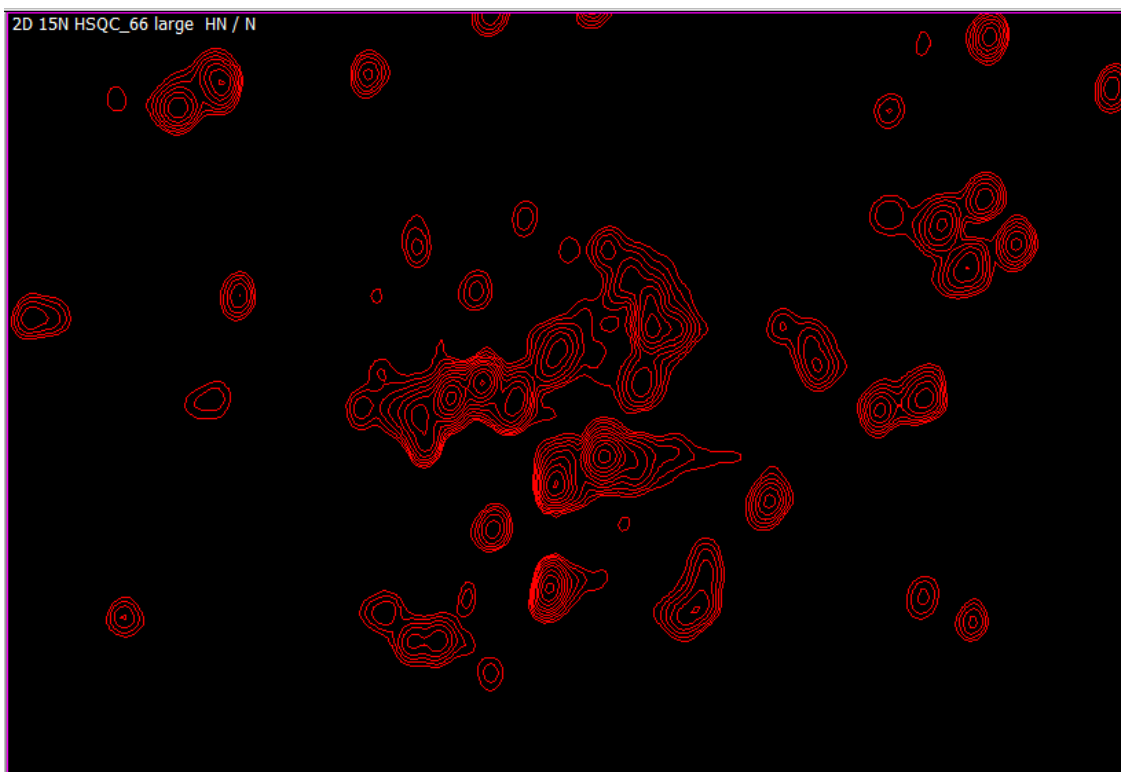
Figure 4.11: ¹⁵N HSQC of Co-MsrB1: Enlargement of the central area of the ¹⁵N HSQC spectrum of Co-MsrB1 displayed in Figure 4.10, showing several overlapping peaks.
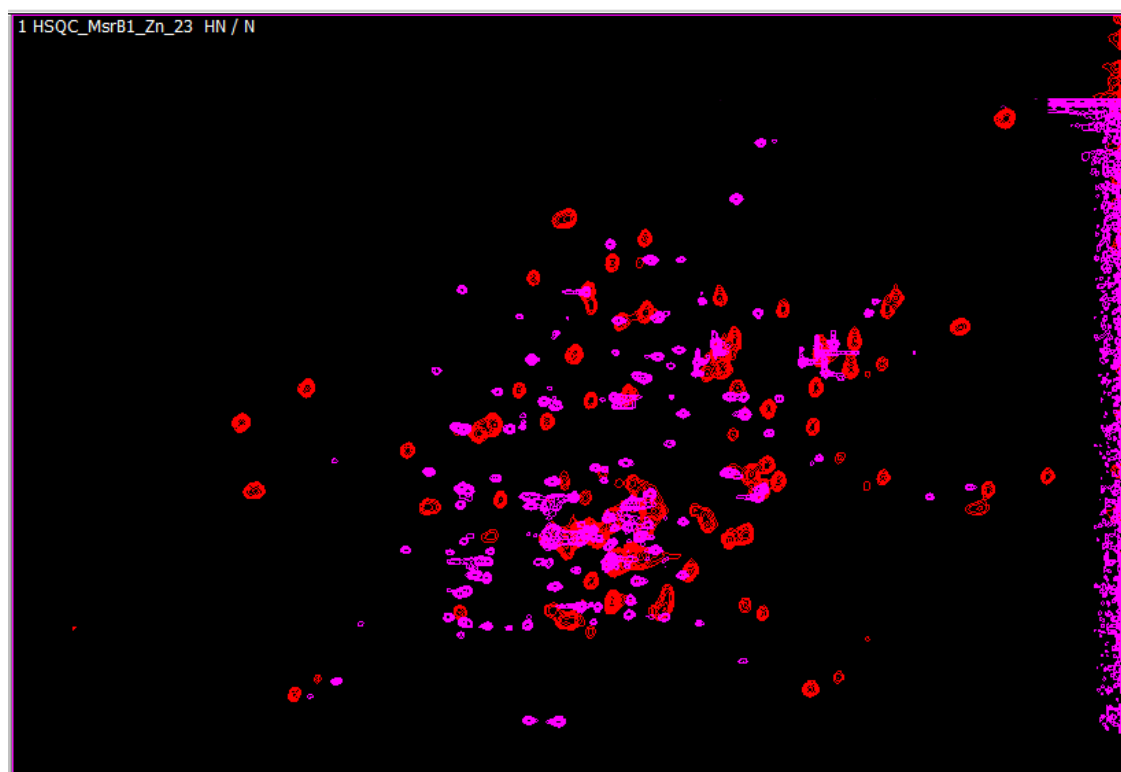
Figure 4.12: ¹⁵N HSQC of Co- and Zn-MsrB1: ¹⁵N HSQC spectra of Co-MsrB1 (red) and Zn-MsrB1 (purple) overlaid. The two spectra are unique.
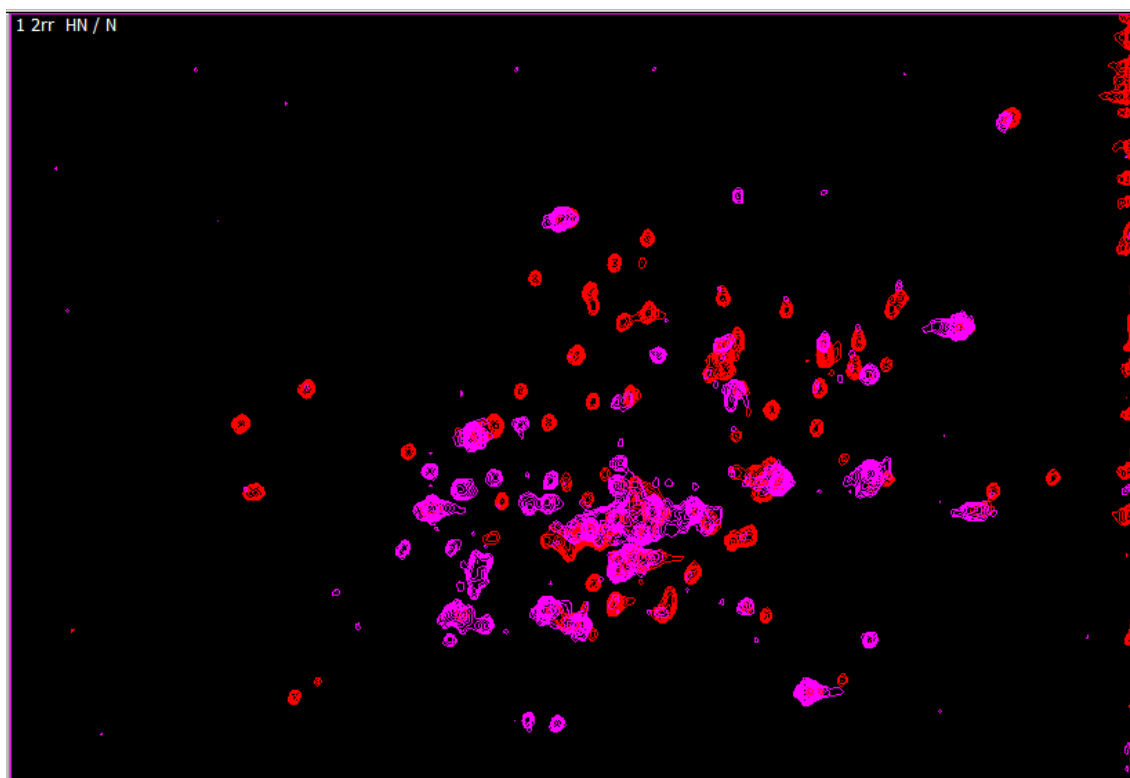
Figure 4.13: ^{15}N HSQC of Co-MsrB1 in 10 and 99.9 % D$_2$O: ^{15}N HSQC spectra of Co-MsrB1 in 10 % D$_2$O(red) and 99.9 % D$_2$O (pink) overlaid. The two spectra have several superimposed cross peaks.
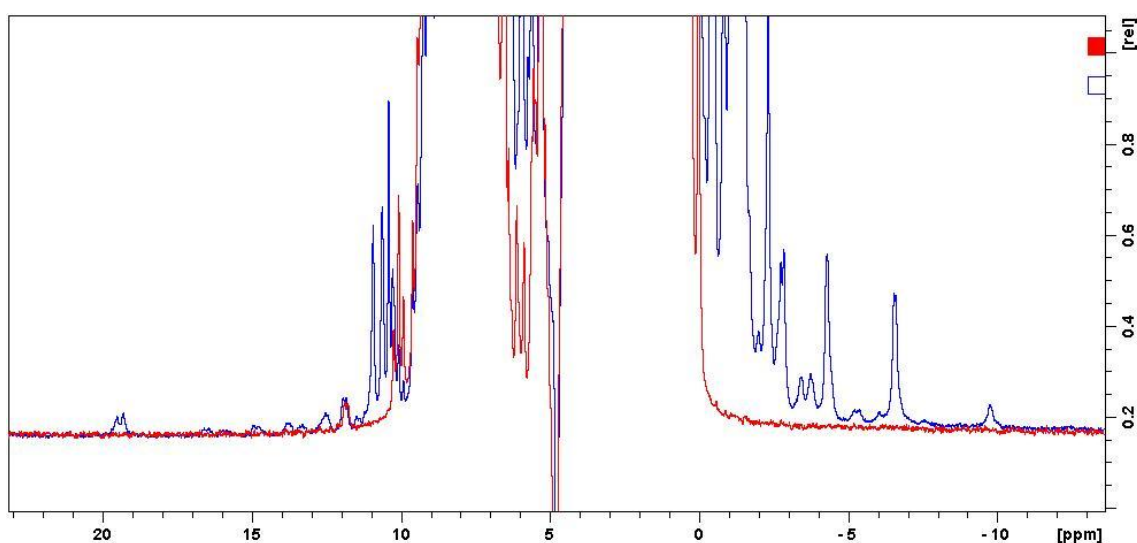
Figure 4.14: 1D ^{1}H NMR spectra of Zn- and Co-MsrB1 (23 to -14 ppm). 1D ^{1}H NMR spectra of Zn-MsrB1 (red) and Co-MsrB1 (blue) (spectral window 23 to -14 ppm) overlayed, showing that several signals are shifted upfield and downfield for Co-MsrB1**.**
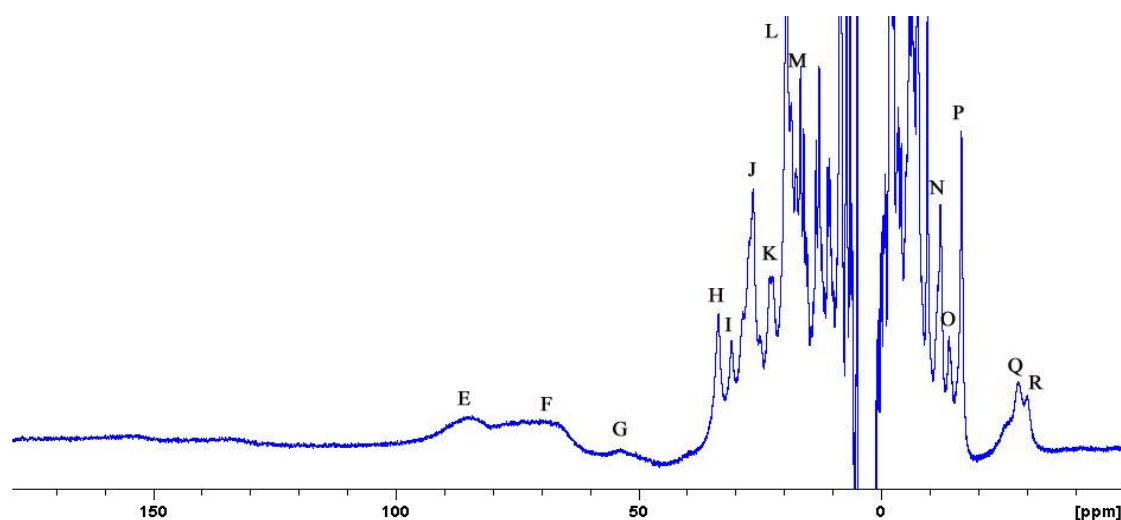
Figure 4.15: 1D ^{1}H NMR spectrum of Co-MsrB1 (180 to -50 ppm). 1D ^{1}H NMR spectrum of Co-MsrB1 (spectral window 180 to -50 ppm). Signals shifted upfield and downfield are labeled with letters.
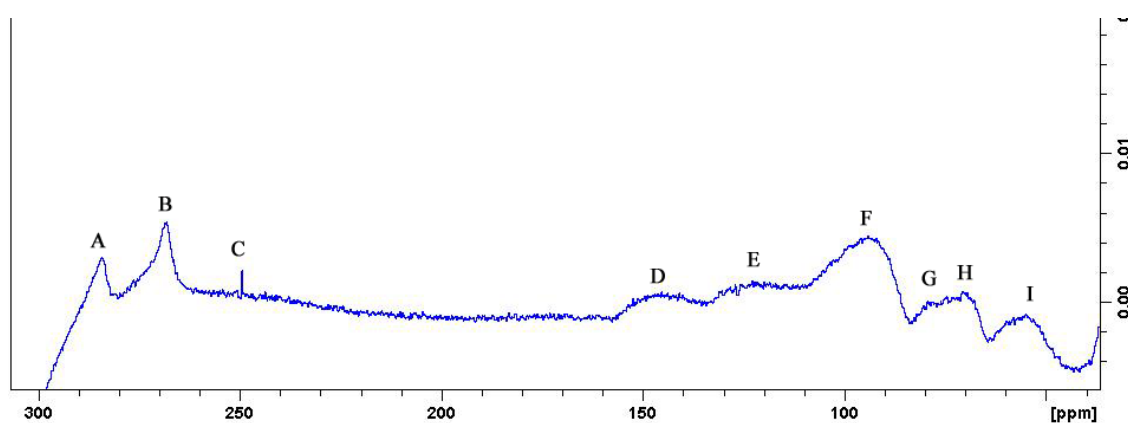
Figure 4.16: 1D ^{1}H NMR spectrum of Co-MsrB1 (300 to -30 ppm). 1D ^{1}H NMR spectrum of Co-MsrB1 (spectral window 300 to -30 ppm) showing signals shifted upfield. Signals are labeled with letters that correspond to the signal labelling of Fig. 4.16. This spectrum is phased differently than 4.16.
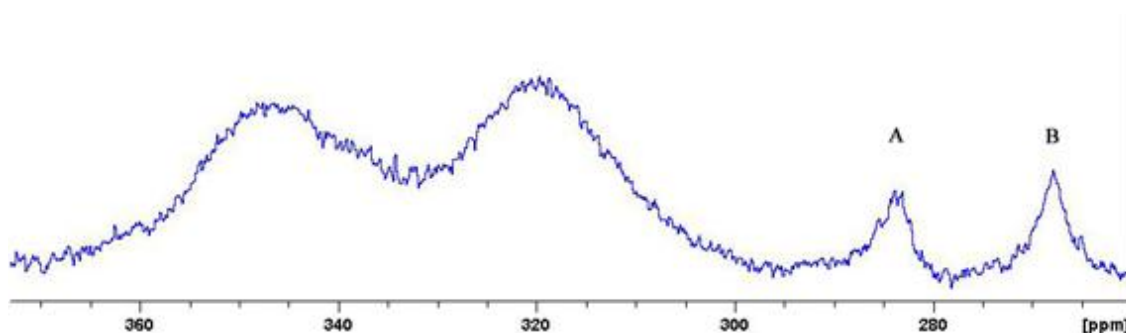
Figure 4.17: 1D ^{1}H NMR spectrum of Co-MsrB1 (370 to 260 ppm). 1D ^{1}H NMR spectrum of Co-MsrB1 (spectral window 370 to 260 ppm). Signals were detected as high as 350 ppm. The two peaks to the right are labelled with letters corresponding to the labelling of Fig. 4.16-4.17.

4.7.2 Assignment of NMR signals

Assignment of NMR signals to backbone and side chain-atoms of Co-MsrB1 was performed as described in section 3.2.7.3.

Assignment were made to backbone atoms (H^N, N, CO, C_α, H_α, C_β and H_β) of 79 of the total 124 amino acids (64 %) that make up His-Co-MsrB1U95C. Of these, Ser98 was previously unassigned for Zn-MsrB1. If Pro- and the His-tag (LEHHHHHH, C-terminal) residues are excluded, backbone assignment of 76 of totally 108 amino acids was achieved (70 %).

Side chain assignment covered 62 of the 79 residues (78 %) for which backbone assignment had been accomplished. Again, this constitutes 50 % of the entire polypeptide (124 amino acids) or 57 % if Pro- and His-tag-residues are excluded.

Crosspeaks were found in ^{15}N HSQC of Co-MsrB1 in D_2O that corresponded to the following peaks: Phe12, Val20, Tyr21, Ala45, Phe46, Ile50, His51, Val55, Lys61, Leu67, Phe82, Phe94, Phe97, Ser100, Leu101, Val104, Ala112 and Leu117.

All assigned backbone- and side chain-atoms are listed together with their chemical shift in Appendix D. The amino acids for which backbone assignment was achieved (including Pro residues) are visualized using the structure for Zn-MsrB1 in Figure 4.18.
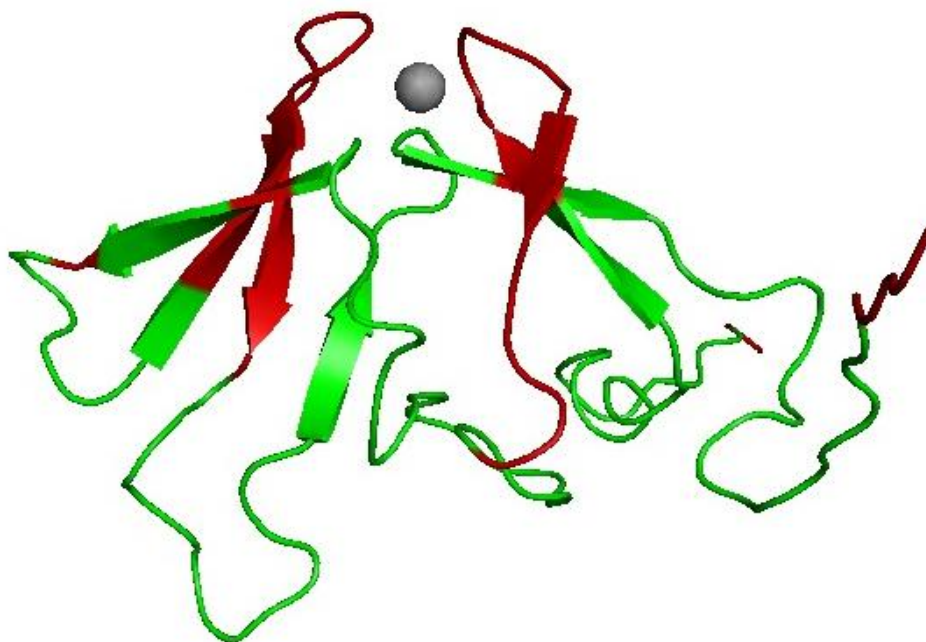
Figure 4.18: Amino acids covered by backbone assignment: The structure of Zn-MsrB1, with backbone-assigned residues (including Pro) of Co-MsrB1 colored green and unassigned residues colored red. The backbone assignment covers 70 % of the polypeptide if Pro- and His-tag residues are excluded.

4.8 Elaboration of NMR data

As described in section 3.2.7.4, the pseudocontact shifts of Co-MsrB1 were used to calculate a magnetic susceptibility tensor for four conformers of the published Zn-MsrB1 structure and used to calculate and predict pseudocontact shifts in the respective conformers for assigned and unassigned residues. For conformers 1, 2 and the mean (average) conformer, the calculated PCSs showed some significant deviations from the observed PCSs, while for conformer 3 these deviations were smaller (Figures 4.19-4.22). The PCSs predicted with AnisoFit for H^N of residues within 7 Å distance to the structural zinc site showed that these residues should have extreme chemical shifts, although the predicted values showed significant variation in magnitude between the conformers (see Table 4.3). Predicted PCSs of H^N for unassigned residues with distance 7-10 Å from the structural zinc binding site in conformer 3 were smaller than residues within 7 Å (see table 4.2).

The data plotted in Figures 4.19-4.22 are listed separately in Appendix E.
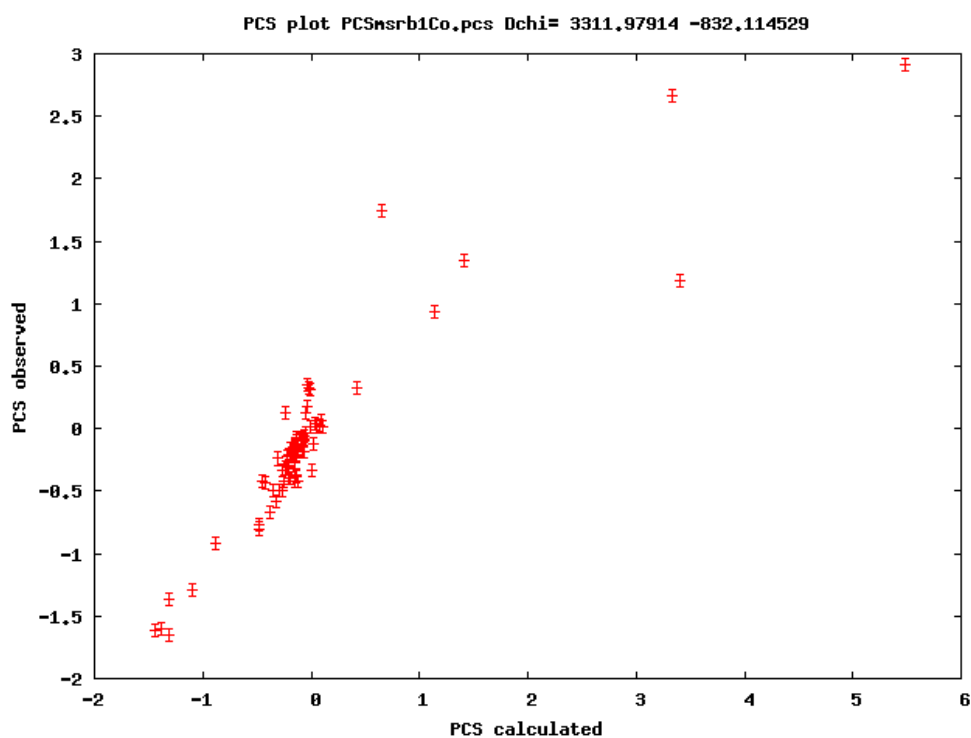
Figure 4.19: Calculated PCS vs. Observed PCS, conformer 1: Observed PCSs vs. PCSs calculated using conformer 1.
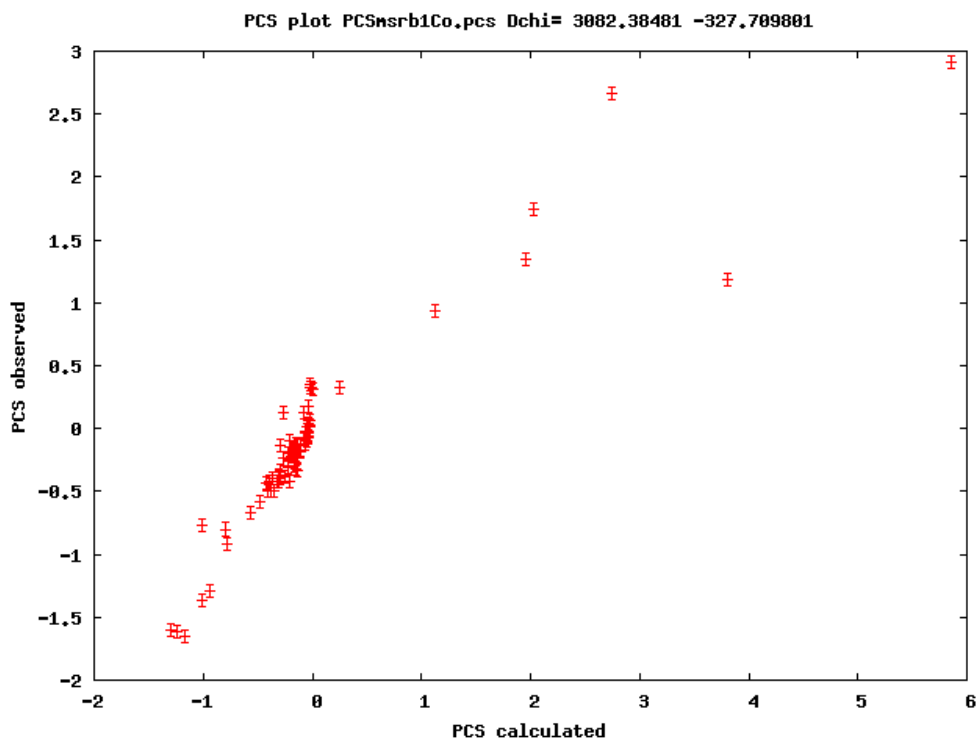
Figure 4.20: Calculated PCS vs. Observed PCS, mean conformer: Observed PCSs vs. PCSs calculated using the mean conformer.
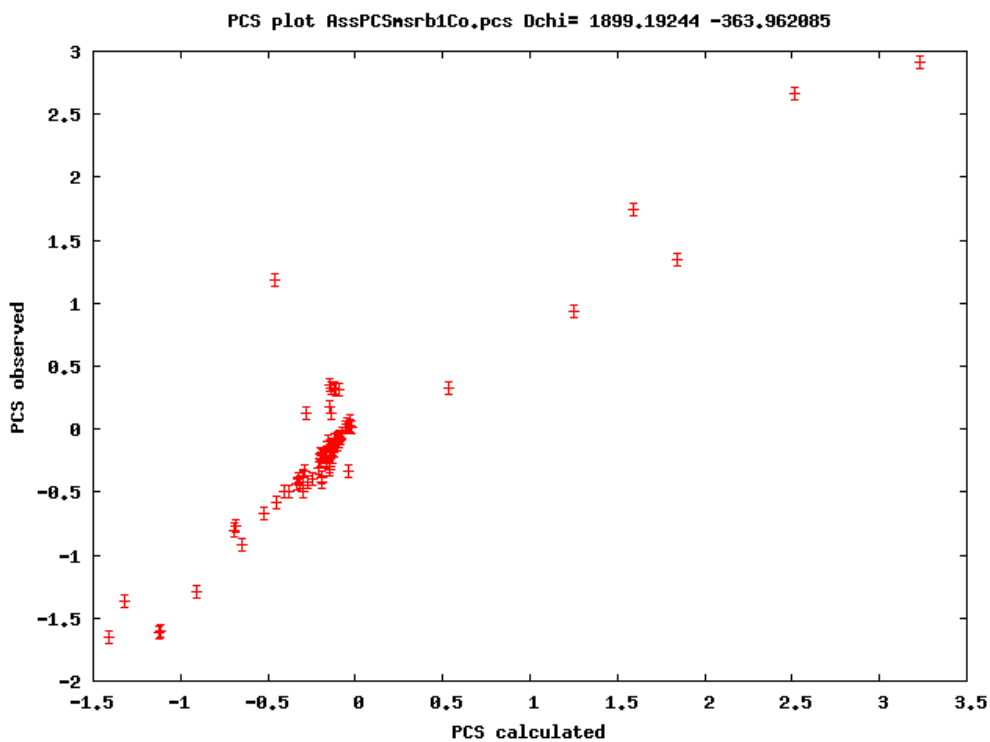
Figure 4.21: Calculated PCS vs. Observed PCS, mean conformer: Observed PCSs vs. PCSs calculated using the mean conformer.
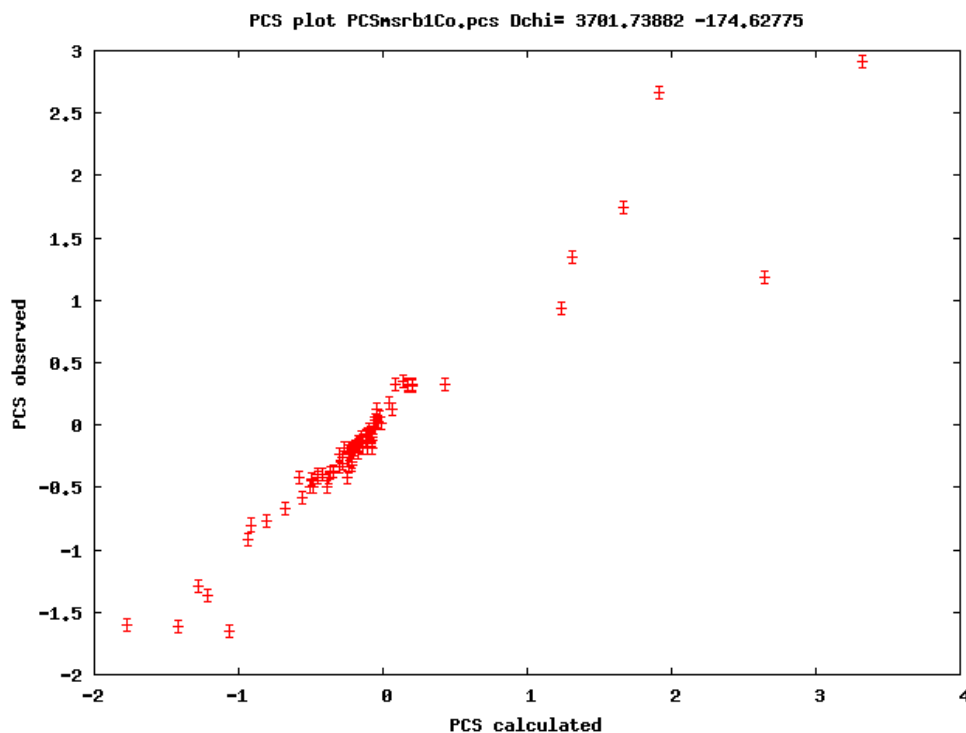
Figure 4.22: Calculated PCS vs. Observed PCS, conformer 3: Observed PCSs vs. PCSs calculated using conformer 3. Calculations using this conformer gave the best correlations between calculated and observed PCSs.

Table 4.3: Precicted PCSs. PSCs predicted for H^N of residues within 7 Å distance to the structural zinc-binding site of MsrB1 (representing the hypothetical Co-binding site) for conformer 1, 2, 3 and the mean conformer (white). PCSs predicted for H^N of residues of 7-10 Å distance of the zinc-binding site for concormer 3 (gray).

Residue	Predicted PCS (ppm), conformer 1	Predicted PCS (ppm), conformer 2	Predicted PCS (ppm), conformer 3	Predicted PCS (ppm), mean conformer
Ala 24	-5,58	-5,67	10,41	-13,69
Lys 25	15,78	14,71	57,31	-12,14
Ser 27	-26,64	-22,2	-24,57	-19,47
Tyr 28	-13,93	-7,31	-16,66	-8,22
Gly 72	-12,58	-7,01	-13,05	-9,56
Lys 73	-28,89	-9,03	-31,78	-23,15
Gly 75	-24	-20,06	-41,95	-14,03
Asn 76	-7,29	-8,86	-10,46	-4,18
GLU 29			-2,25	
LEU 30			-2,5	
THR 56			-4,28	
SER 70			-3,68	
GLY 77			1,6	
LYS 102			-3,46	

Plots of PCSs for H^N vs. hypothetical distance to Co were made as described in section 3.2.7.4.

For all plots using H^N to Zn-distances as calculated from the published conformers, PCSs for H^N were inversely proportional to distance with the exception of 10 residues (see Figures 4.23-4.25):

- T49, I50, H51, D53, S54, V55, F97, S99, S100, L101

A plot of PCSs of H^N vs. distance using the C' atom of Lys25 as a hypothetical placement of the cobalt metal ion showed a significantly impaired PCS-distance relationship (Figure 4.26). A plot of PCSs of H^N vs. distance to His120 using conformer 1 showed no proportionality between distance and PCS (Figure 4.27).

Chemical shifts of H^N for Co- and Zn-MsrB1, the resulting PCSs and calculated H^N-Zn distance is shown for conformer 1 in Appendix F as an example of data used to create the plots shown in Figures 4.23-4.27.
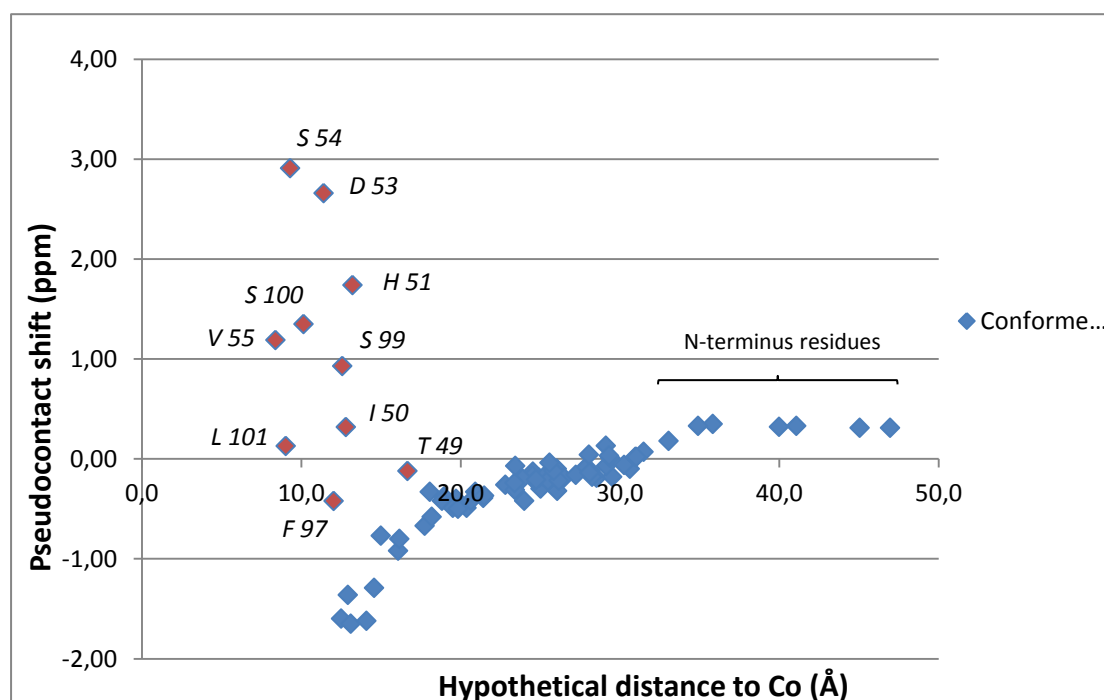
Figure 4.23: Plot of PCSs for H^N vs hypothetical distance to Co: PCS for H^N for each residue that was found in the backbone assignment plotted against the distance of the individual H^N to the Zn-ion in conformer 1 of the Zn-MsrB1 structure. H^N that follow the inverse proportionality of PCS and distance are shown in blue and H^N that deviate from this trend are shown in red with a label of the residues to which they belong. H^N that belong to the end of the N-terminus are indicated by a brace.
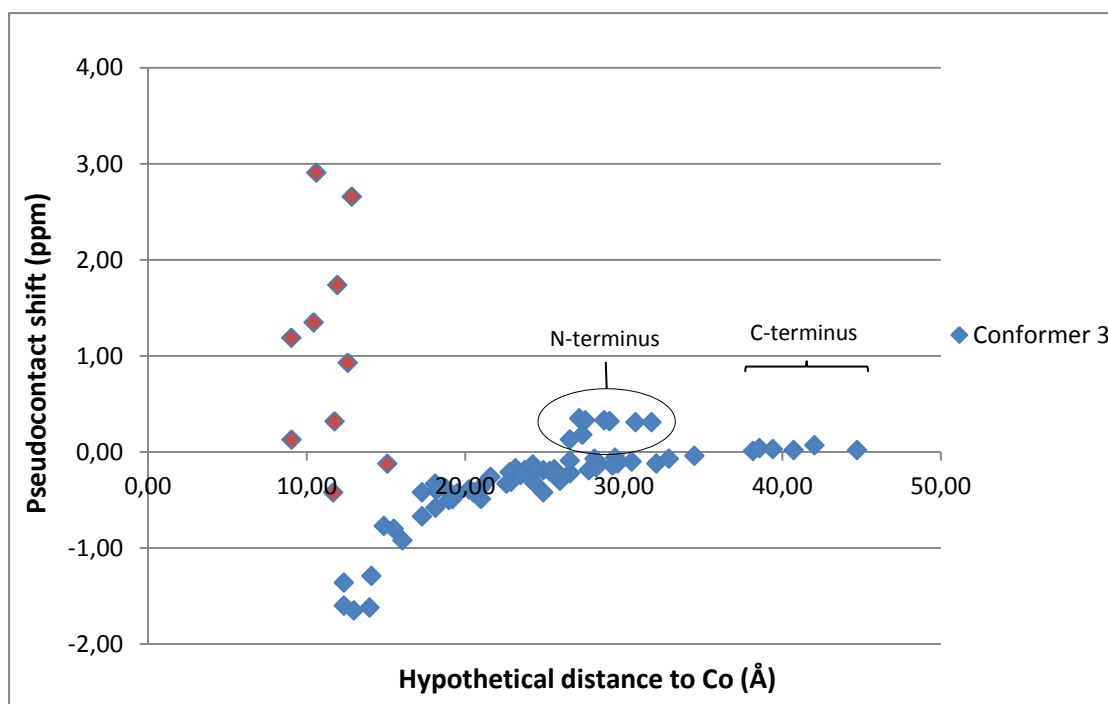
Figure 4.24: Plot of PCSs for H^N vs hypothetical distance to Co: PCS for H^N for each residue that was found in the backbone assignment plotted against the distance of the individual H^N to the Zn-ion in conformer 3 of the Zn-MsrB1 structure. H^N that follow an inverse proportionality of PCS with distance are shown in blue and H^N that deviate from this trend are shown in red (the same residues as in Figure 4.23). H^N that belong to the end of the N-terminus are indicated by a brace.
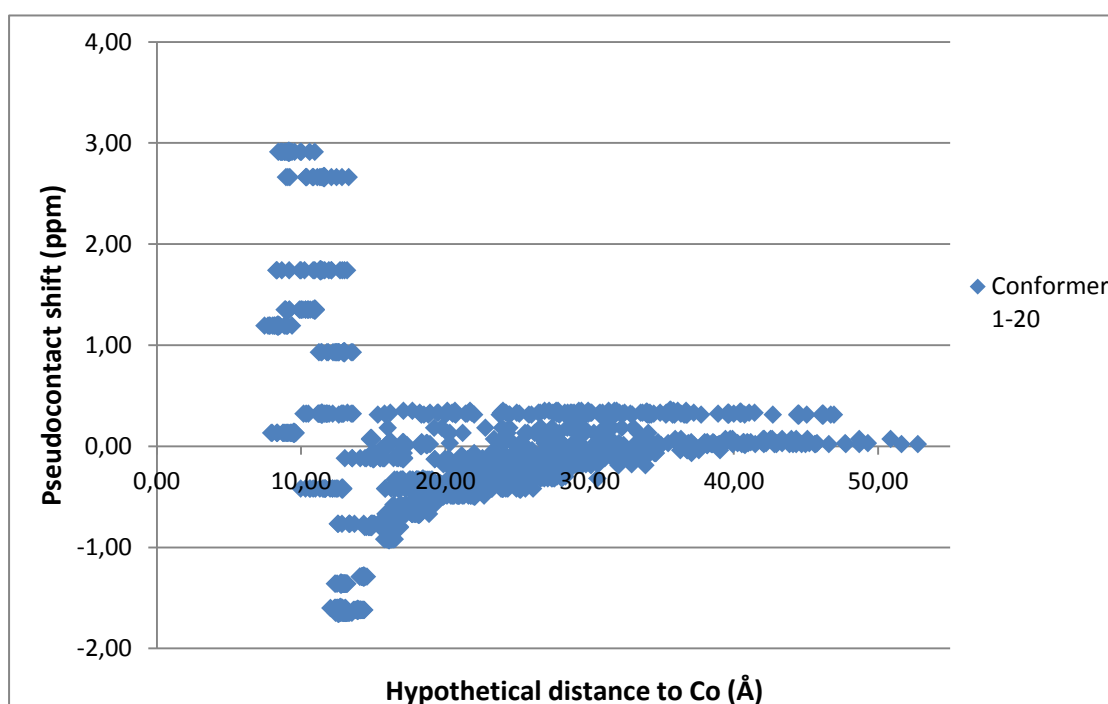
Figure 4.25: Plot of PCSs for H^N vs hypothetical distance to Co: PCS for H^N for each residue that was found in the backbone assignment plotted against the distance of the individual H^N to the Zn-ion for as calculated for conformers 1-20 of the Zn-MsrB1 structure (hypothetically representing the Co-ion and the structures of Co-MsrB1, respectively).
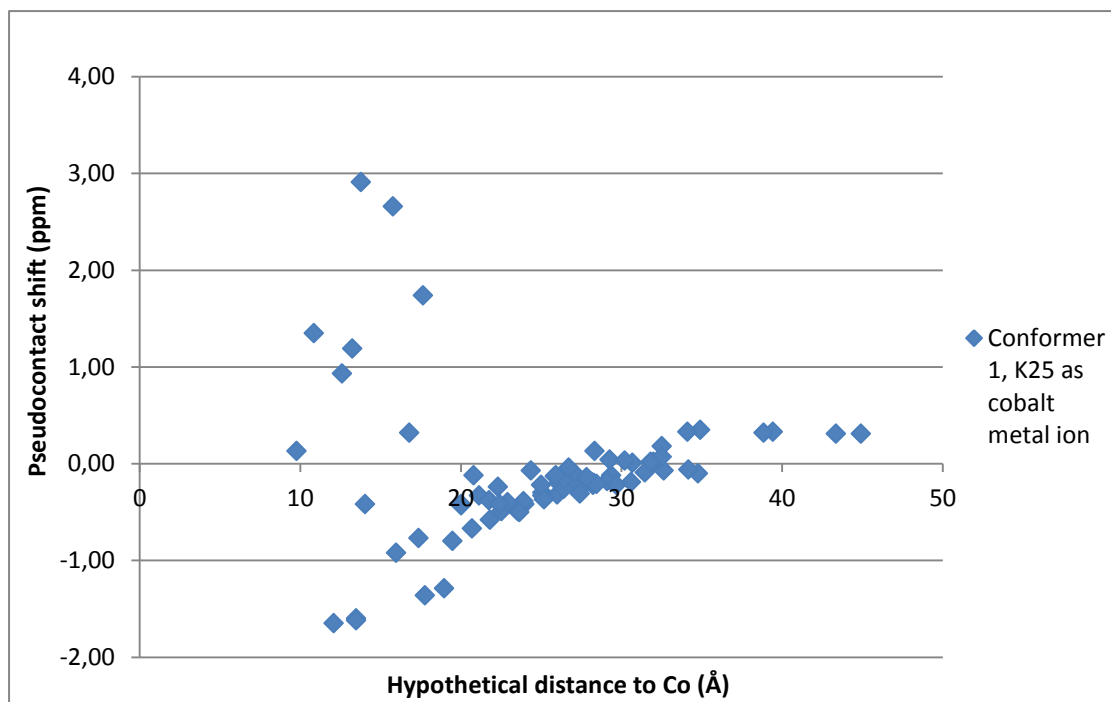
Figure 4.26: Plot of PCSs for H^N vs hypothetical distance to Co: PCS for H^N for each residue that was found in the backbone assignment plotted against the distance of the individual H^N to C' atom of K25 in conformer 1 of the Zn-MsrB1 structure (hypothetically representing the Co-ion the structure of Co-MsrB1, respectively.
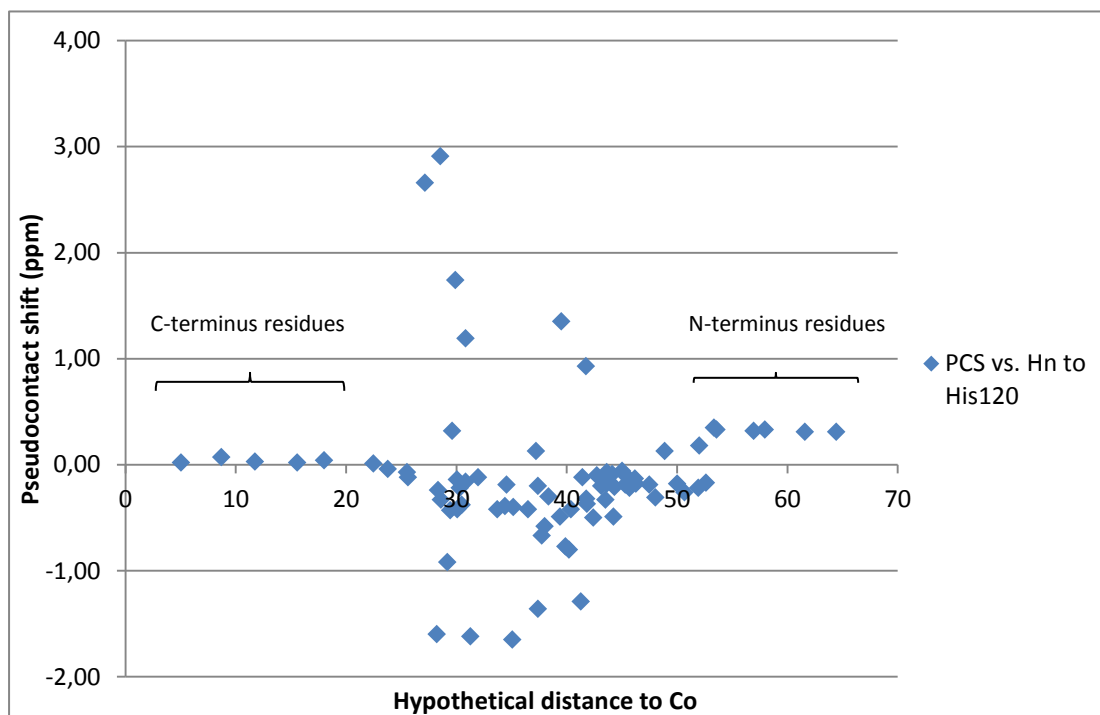
Figure 4.27: Plot of PCSs for H^N vs hypothetical distance to Co: PCS for H^N for each residue that was found in the backbone assignment plotted against the distance of the individual H^N to His120 in conformer 1 of the Zn-MsrB1 structure (hypothetically representing the Co-ion bound to the His-tag and the structure of Co-MsrB1, respectively). H^N belonging to the ends of the C- and N-terminus are indicated by separate braces.

5 Discussion

5.1 Effects of cobalt(II) on culture growth

In this study, *E. Coli* was grown in lysogeny broth (LB) to produce Zn-MsrB1 and in cobalt-supplemented media to produce Co-MsrB1. As shown in section 4.1, the inhibitory effects of cobalt were manifested as decreasing growth rates with increasing concentration of $CoCl_2$. Although there was very little difference in the growth rates between M9 medium with 2 and 10 µM $CoCl_2$, indicating that lower concentrations of cobalt do not contribute significantly to cellular stress, growth rates were significantly higher in LB than for M9 with 2 of 10 µM $CoCl_2$. Indeed, in contrast to LB, which is a rich medium based on yeast extract and peptone (and thus containing all amino acids, nucleotides, fatty acids, etc.) M9 medium is only supplied with sugar and ammonium as carbon- and nitrogen source, respectively. The energy spent on biosynthesis in M9 medium should therefore be dramatically higher, leading to impaired growth.

For 50-150 µM $CoCl_2$, toxic effects are clearly observable after 2 hours with about half the measured OD_{600} compared to 2 and 10 µM $CoCl_2$. From this point growth slows down and the concentration differences between the media become more apparent. Finally a stable, linear growth rate is reached. This shows that there is an initial buildup of intracellular cobalt before the detection of cobalt by metallosensory regulatory proteins leads to upregulation of the cobalt/nickel efflux pump and downregulation of Fe-S-cluster proteins [14]. It can then be suggested that the linear growth rate then represents a state where the intracellular concentration of cobalt has reached a plateau, with full activation of the cobalt efflux- and adaption systems.

For production of Co-MsrB1, *E. coli* was first starved of zinc before inoculation into zinc-free cobalt-supplemented M9-medium. Despite the zinc-starvation, the bacteria exhibited logarithmic growth. This was somewhat surprising as zinc is required by *E. coli* in many proteins, and indicates that *E. Coli* may utilize other metals as a substitution for zinc in these proteins (see Further studies, section 5.10).

5.2 Protein expression and purification of MsrB1

The mammalian protein MsrB1 was expressed using a bacterial vector. Eukaryotic proteins can be difficult to express with bacterial vectors as it often leaves the protein in insoluble inclusion bodies [45] or gives very low yield or incorrect folding and/or aggregation of the protein [46]. The MsrB1 expression vector had already been used in previous work and optimization of this vector was therefore not necessary as it consistently gave high protein yield and purity, and MsrB1 was also obtained with high purity in this study as shown in section 4.2. Although some unspecific protein was sometimes visible in SDS-PAGE of the elution fraction, these bands were consistently about 30 kDa or larger. As proteins with a high molecular weight will give larger bands than the same number of proteins with low molecular weight, the amount of unspecific protein compared to MsrB1 is exaggerated in SDS-PAGE. MsrB1 is a relatively small protein, and the His-tag is at the end of a long and flexible C-terminus. This is likely to have contributed to the high yield and successful purification with Ni-NTA affinity matrix. Purified MsrB1 also showed high stability in solution, making it a convenient protein to work with. MsrB1 belongs to the nucleus and cytosol *in vivo* and has a structural metal binding site, making it naturally water soluble and less prone to denaturation.

5.3 Metal-uptake by MsrB1

As demonstrated in section 4.3, uptake of Co by MsrB1 was demonstrated for all Co-supplemented media. It has previously been demonstrated that the His-tag is not responsible for this uptake (Dobrovolska et al., unpublished results). As can be seen of Figures 4.3 and 4.4, five distinct UV-vis absorption peaks were observed for Co-MsrB1 at 335, 631, 660, 695 and 723 nm. Absorption at 335 nm is typical for cobalt(II) ion-S⁻-Cys charge transfer, while absorption at 631, 660, 695 and 723 nm represent d-d transitions of high-spin cobalt (II) d-electrons [47, 48]. The 631, 660, 695 and 723 nm are characteristic of tetrahedral four-Cys-coordination of cobalt(II) [47, 48].

MsrB1 has seven Cys-residues: four of which are responsible for zinc-coordination in the native protein (Cys 23,24, 71 and 74) two of which are involved in the catalytic mechanism (Cys 4 and 95) and one which does not have a specific structural or catalytic function (Cys 58). Enzyme activity measurements show that Co- and Zn-MsrB1 have similar catalytic activity (Dobrovolska et al., unpublished results), which indicate that the Cys4 and Cys95 cannot be involved in coordination of cobalt. Hypothetically, this leaves five remaining Cys residues as possible ligands for cobalt. However, since Co-MsrB1 has enzymatic activity, it can be assumed that the overall structure of the catalytic site is similar to that of Zn-MsrB1. Further, since the catalytic Cys resides in a β-sheet this suggests an overall similar fold of Co-MsrB1. Cys58, Cys71 and Cys73 also reside on this β-sheet (Figure 5.1). In a scenario where Cys58 is involved in tetrahedral cobalt coordination together with three Cys residues from the zinc-binding site, this β-sheet should be impossible. These results therefore suggest that cobalt is bound to the same Cys residues that make up the zinc-binding site in the native
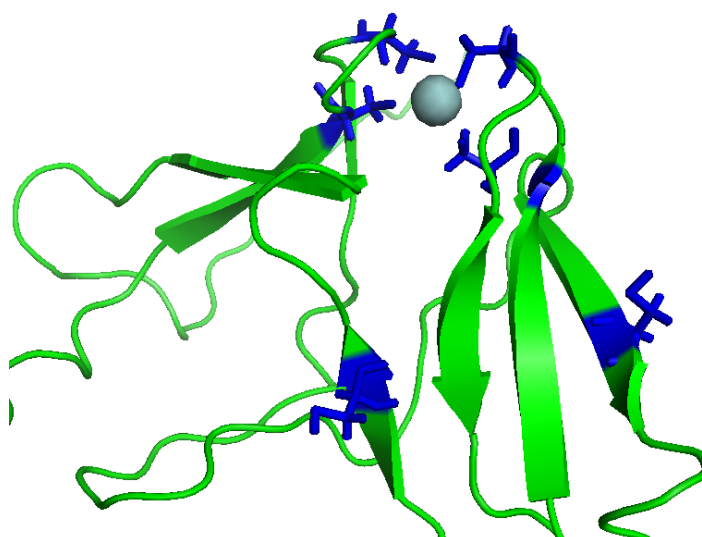
Figure 5.1: Cysteines of MsrB1: Six of seven cysteine residues of MsrB1 as they reside in the structure of Zn-MsrB1, the backbone and side chains shown in blue, with side chains visualized as sticks. The catalytic Cys95 (bottom, left) and Cys58 (right) reside on the same β-sheet. Coordination of cobalt by Cys58 would ruin the β-sheet and subsequently the active site.

protein, replacing zinc as a structural metal ion.

E. Coli was grown in LB supplemented with 10 µM NiCl$_2$ in an attempt to produce Ni-MsrB1, as described in section 3.2.1.4, and this experiment was performed to evaluate if the *in vivo* substitution of zinc for cobalt in MsrB1 represents a unique quality of this metal binding site towards cobalt or if other metal-substitutions could be achieved *in vivo*. Although little literature is available on UV-vis absorption spectra for tetrahedral amino acid-coordination of nickel, characteristic absorption appears to be in the 400-500 nm range [49], yet no such absorption peaks were detected. Uptake of cobalt by MsrB1 in the presence of zinc violates the Irving-Williams series. In addition, MsrB1 does not bind nickel under the same conditions of which MsrB1 binds cobalt, which is also inconsistent with the Irving-Williams series. As described in section 1.2.7, *E. Coli* creates a metal ion concentration gradient in the cytosol and periplasm, decreasing and increasing with the Irving-Williams series, respectively. As can be seen in Figure 1.14, the dissociation constant of the zinc sensory protein (which regulates zinc efflux) is in the magnitude of 10^{-12} M, which means that the zinc concentration above 10^{-12} M keeps the zinc efflux protein active. Conversely, the dissociation constant of the cobalt sensory protein is in the magnitude of 10^{-7} M [15]. The intracellular concentration of cobalt should thus in principle be 100,000 times higher than for zinc. Furthermore, it has been shown that one protein that binds both manganese and copper (two metals at opposite ends of the Irving-Williams series) binds manganese *in vitro* when manganese is present in 10,000-fold excess to copper. Cobalt and zinc are adjacent in the Irving-Williams series, and the high concentration of cobalt therefore likely compensates for its lower affinity in the competition for MsrB1 Cys-ligands, leading to formation of Co-MsrB1. In addition, the structure of zinc-binding sites are optimized for the ionic radius of zinc, and the ionic radius of zinc(II) in tetrahedral coordination is closer to cobalt(II) than any other metal in the Irving-Williams series (0.60 and 0.58 Å, respectively) [4], which should favor substitution of cobalt over the other metals. As described in section 1.2.7, nickel is not released directly into the cytosol but delivered to the correct polypeptides by metallochaperones, and explains why no nickel uptake by MsrB1 was detected. It is therefore argued that the metal uptake characteristics of MsrB1 demonstrated in this study is due to both the intracellular conditions created by *E. Coli* and the very similar ionic radii of zinc(II) and cobalt(II).

5.4 pH-dependence of the molar extinction coefficient for Co-MsrB1

The molar extinction coefficient (ε) was found for Co-MsrB1 in solutions with pH-values ranging from 4.9 and 11.5, and the plot of these in Figure 4.5 identifies two possible titration curves: from 4.9 to 6.3 and 7.2 to 10.2. pK_a represents the pH at 50 % protonation of the side chain [50]. The first titration curve gives a pK_a of approximately 5.6 (Figure 5.2), while the second curve should represent two titrations from 7.2 to 9.2 and 9.4 to 10.9 and thus give pK_as of approximately 8.0 and 10.1, respectively (Figure 5.3).
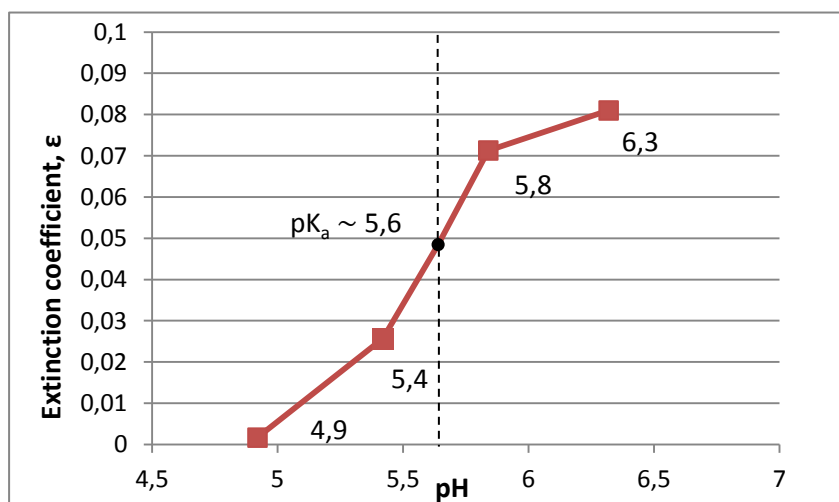
Figure 5.2: pH-titration curve 1. ε of Co-MsrB1 for pH 4.9-6.3 at 695 nm. The titration-curve indicates a pK_a of approximately 5.6, indicated with a stipulated line and a black marker as it intersects with the curve. The pH of the four titration points are labeled next to the data points.
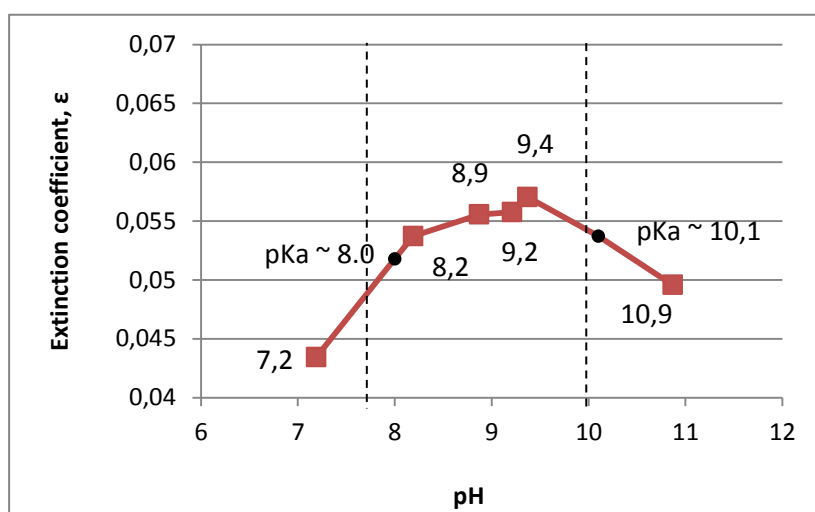
Figure 5.3: pH-titration curve 2. ε of Co-MsrB1 for pH 7.2-10.9 at 695 nm. The titration-curve indicates two pK_as of approximately 8.0 and 10.1, indicated with a stipulated line and a black marker as they intersect with the curve. The pH titration points are labeled next to the data points.

As the three characteristic absorption peaks used to measure ε represent d-d transitions of the cobalt metal ion, the titration curves could immediately be interpreted as protonation of the cobalt-coordinating Cys-residues. However, protonation of these residues should lead either to reduction in the number of ligands, for which there would be a shift in the absorption peaks to lower wavelengths [25], or complete dissociation of the complex. As can be seen in Appendix C, no absorption shift was observed in any pH, showing that absorption is caused by a four-coordinated cobalt-Cys complex. The titration curves are therefore interpreted as either caused by different degrees of dissociation of the cobalt-Cys complex or pH titrations of amino acid side chains in the vicinity of the cobalt-Cys complex. Four residues within reasonable distance have titratable side chains and thus are possible candidates: Lys25, His51, Lys73 and Asn76.

Although these experiments indicate that there are titratable residues, further investigation is needed before a proper conclusion can be made from these results (see Further studies, section 5.10).

5.5 Metal removal from Co- and Zn-MsrB1 and reconstitution

As shown in section 4.5, removal of Co from Co-MsrB1 was achieved overnight at pH 5.0 and over the course of six days for pH 5.5, but not for pH 7.0 and 7.5. Removal of Zn-MsrB1 was not detected at any pH. This shows that the Zn-Cys bonds are stronger than Co-Cys. This is consistent with the Irving-Williams series, which predicts the zinc-complex to be more stable. Furthermore, the results of these experiments show that metal chelating agents alone are not able to remove the metal ion of both Zn- and Co-MsrB1, indicating that the metal ion of both proteins is tightly bound to the metal binding site and/or is inaccessible to the chelating agents.

Apo-MsrB1 was dialyzed against buffers of different pH and ionic strength containing $CoCl_2$ in order to reconstitute Co-MsrB1, and uptake of cobalt by apo-MsrB1 was not detected by UV-vis spectroscopy or 1D ^{1}H NMR. As formation of metal-bound protein is assumed to result from spontaneous self-assembly *in vivo*, it should in principle be possible to reconstitute Co-MsrB1 by increasing the pH above the pK_a of the coordinating Cys residues and supplying Co(II) to the solution. However, several challenges appeared in these experiments. At pH 5 and 5.5, Co(II) formed strongly colored complexes with sulfide-group reducing agents like β-mercaptoethanol and DTT. This makes UV-vis absorption of the protein solution useless and lowers the availability of Co(II) for the apo-protein. At pH 7 and 7.5, Co(II) precipitated as $Co(OH)_2$ even at lower concentrations, forming a pink solid that covered the beakers and inside of the dialysis membrane. In addition, apo-MsrB1 precipitated after overnight dialysis at pH 7 and 7.5, eventually leading to a complete loss of dissolved protein. Successful reconstitution with thiolate ligands may need strict anaerobic conditions [51], a requirement which was not met in the experiments of this study. In the attempts at reconstitution relatively high concentrations of $CoCl_2$ were also used, something which is not necessary [51] and as mentioned led to precipitation of $Co(OH)_2$.

5.6 Determination of Co-MsrB1:Zn-MsrB1 ratios in mixed Co- and Zn-MsrB1 samples

As shown in section 4.6, the Co-MsrB1:Zn-MsrB1 ratios for M9-medium supplemented with 10 µM $CoCl_2$ and 10 µM $ZnSO_4$,LB supplemented with 10 $CoCl_2$ and LB supplemented with 50 µM were found to be about 0.2 (20 %), 0.1 (10 %) and 0.03 (3 %), respectively.

Metal uptake by a newly formed polypeptide chain to form a metalloprotein is regarded to be a process of spontaneous self-assembly. The uptake of cobalt by MsrB1 in the presence of zinc, violating the Irving-Williams series, was discussed in section 5.3. A higher Co-MsrB1:Zn-MsrB1 ratio in M9 medium than LB-medium indicate that the intracellular concentrations of cobalt are higher in the M9 medium or that the intracellular zinc concentration is higher in the LB media. As discussed in section 5.1, the M9 medium represents a stress condition to *E. coli* in which more energy is spent on biosynthesis. As *E. Coli* has different influx-, sensory- and efflux systems for zinc and cobalt [14, 52, 53], it is argued that the cobalt sensory- and efflux system or zinc influx system somehow suffer under the stress condition of M9 medium, leading to a higher intracellular concentration of cobalt relative to zinc compared to *E. Coli* grown in LB medium.

The quality of the method behind the ratio calculations should also be addressed. Although the extinction coefficient (ε) was shown to be (near) constant for the Co-MsrB1 solutions with different concentration, the amount of Co-MsrB1 in the protein sample from LB-medium supplemented with 10 µM $CoCl_2$ was measured to about 3 %. The UV-vis absorption spectra have a resolution of three digits, and absorption at 695 and 723 nm was 0.018 and 0.016, respectively. This indicates that precision of this particular measurement is not as good as for the samples with higher concentration, and that this protein sample should have been concentrated further before recording the absorption spectrum. Overall, however, the method shows a low standard deviation (less than 5 %), indicating that the method has high precision. The calculated ratios of the respective samples should thus be a good representation of the relative ratio-difference between the three growth media. Integration of peaks of a ^{15}N HSQC NMR spectrum of a mixed sample may have provided more precisely determined Co-MsrB1:Zn-MsrB1 ratios. However, determination by UV-vis absorption spectroscopy is inexpensive since no isotopic labeling is required and ratios can thus be determined for more samples.

5.7 NMR assignment of Co-MsrB1

2D- and 3D NMR spectra of Co-MsrB1 were recorded and backbone- and side chain-assignment was performed and the results reported in section 4.7. Of the polypeptide backbone, 79 out of 124 amino acids (64 %) of were assigned using the ^{15}N HSQC and triple resonance spectra or 76 of 108 (70 %) if Pro- and His-tag residues are disregarded. Since Pro-residues do not have H^N atoms and thus cannot be seen in the spectra, and the His-tag can be difficult to assign due to the six consecutive His-residues that are not part of the native polypeptide, it is reasonable to disregard these residues. This leaves in total 32 residues unassigned for Co-MsrB1. As can be seen from Figure 4.18 and in Appendix F, the four Cys-residues that make up the zinc-binding site of the native protein are unassigned, whereas the other three Cys-residues of Co-MsrB1 were assigned. Nuclei of Cys-residues that are bound to a paramagnetic metal ions experience both contact and pseudocontact shift contribution, which result in extreme chemical shifts of several hundred ppm and extreme broadening of the signals of nearby nuclei (up to several hertz) depending on the paramagnetic metal ion. These results therefore confirm that the same Cys-residues are responsible for metal ion-coordination in both Co- and Zn-MsrB1. As can be seen from figures 4.16-4.17, several signals were observed between 50 and 350 ppm in 1D ^{1}H spectra of Co-MsrB1 that should thus belong to ^{1}H nuclei of the cobalt-coordinating Cys-residues (the most extreme shifts of which should then belong to H_βs of these Cys-residues as they are closest to cobalt(II)). Furthermore, the unassigned residues constitute two stretches that each include two of the cobalt-coordinating Cys-residues. Residues adjacent or nearby to the cobalt-coordinating Cys-residues also experience strong pseudocontact shifts due to the short distance to the cobalt ion. As can be seen from Figures 4.14 and 4.15, many signals were observed above 15 ppm and below 5 ppm (outside the spectral width of the processed ^{15}N HSQC spectrum), and the majority of unassigned residues should thus be found here (see section 5.8 for further discussion).

Side chains of 62 of 124 residues (50 %) of the entire polypeptide or 59 of 108 (57 %) residues if Pro- and His-tag residues are disregarded were assigned using the TOCSY spectrum of Co-MsrB1. This constitutes 78 % of the backbone-assigned residues. Consistent with the results of the backbone assignment, the majority of residues which could not be assigned were residues considered to be close to the cobalt binding site. In the backbone assignment, many spectra were are used in concert, and absence of signals for a residue in one spectrum can be compensated by signals in other spectra (e.g. C_α gives a crosspeak in both CACBNH and NHCA) or signals from the adjacent amino acid, while in TOCSY such oportunities are absent. In addition, ^{1}H nuclei are more sensitive to paramagnetic influence than ^{13}C and [15]N [21]. Together these observations explain why less assignments could be made in TOCSY.

Analysis of the ^{15}N HSQC of Co-MsrB1 in D_2O overlaid with the ^{15}N HSQC spectrum of Co-MsrB1 in 10 % D_2O identified 18 residues with H^N with very slow solvent exchangability. Lack of exchangability with deuterium is generally an indication that the H^N are involved in a

strong hydrogen bond, typical for residues which are part of a tightly packed hydrophobic core of the protein [54]. Consistent with this, all residues identified have hydrophobic side chains.

The 2D- and 3D-spectra of Co-MsrB1 will also be briefly addressed. The ^{15}N HSQC spectrum of Co-MsrB1 showed that a trace of Zn-MsrB1 was present in the sample upon recording of the spectrum, manifested as weaker signals corresponding to those of the published assignment of Zn-MsrB1. These signals were also found in the triple resonance spectra. In total, about 200 crosspeaks were found in the HSQC spectrum (Figure 5.3).
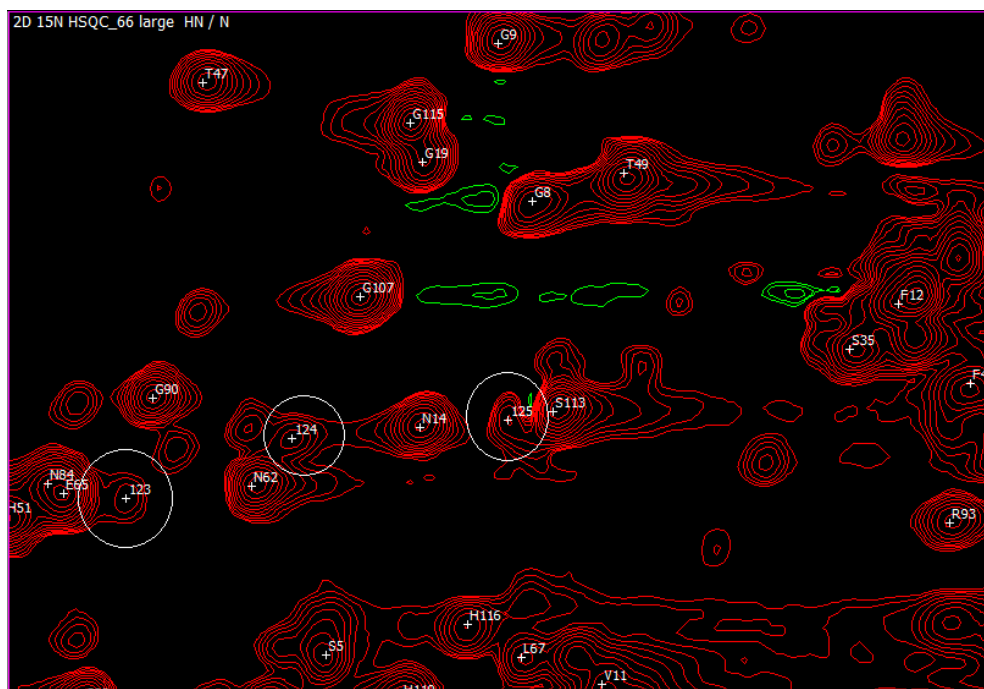
Figure 5.3: Signals from Zn-MsrB1 in ^{15}N HSQC. ^{15}N HSQC of Co-MsrB1 showing three examples of signals from Zn-MsrB1 marked with a number and circle. Assigned crosspeaks of Co-MsrB1 are marked with oneletter code and number corresponding to residue type and –number, respectively. Peak 123 corresponds to N62, 124 to N14 and 125 to S113.

Signals from Zn-MsrB1 served as confirmation of the assignments of Co-MsrB1 when both were identified for a given residue, and was thus sometimes helpful. On the downside, the many signals present in the spectrum led to many overlapping peaks, which made the spectrum time consuming to work with.

5.8 NMR data elaboration and interpretation

A magnetic susceptibility tensor and pseudocontact shifts for assigned and unassigned residues were calculated, and the results were described in section 4.8. As can be seen in figures 4.19-22, calculated and observed PCSs for some residues show rather significant deviations. However, there are also significant differences in these deviations between the four structural conformers used in this analysis. As it was discussed in section 5.3, one can assume that Co- and Zn-MsrB1 have very similar structure, and the differences between the calculations indicate that minor structural misrepresentations have large contributions to the quality of the calculations. In fact, the published structure of MsrB1 has an RMSD for the backbone-atoms (deviation of backbone atoms of the conformer ensemble from the average calculated position) of 1.59 Å [42]. In comparison, a plot of observed vs. calculated PCSs using Fantasia (predecessor of AnisoFit) for a protein structure with an RMSD of 0.7 Å showed large deviations (Figure 5.4).
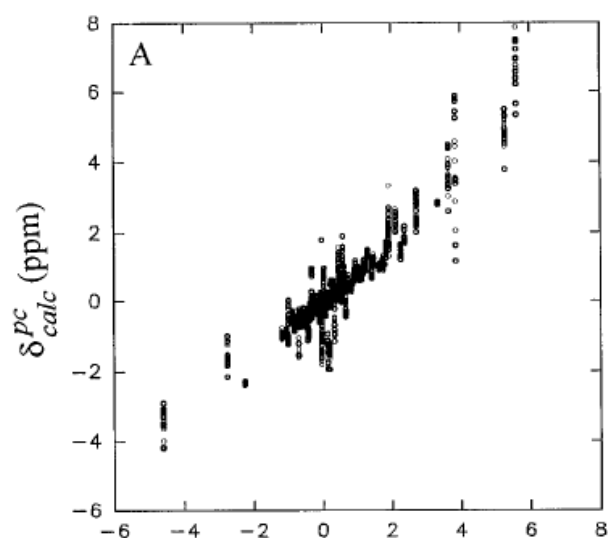
Figure 5.4: Observed PCSs vs. calculated PCSs for Met80Ala cyano-cytochrome c. Observed PCSs vs. calculated PCSs for Met80Ala cyano-cytochrome c, obtained by Fantasia, with calculations for 17 conformers overlaid. The conformers used in the calculations had an RMSD for backbone-atoms of 0.7 Å from the average coordinates, while the published structure of MsrB1 has an RMSD for the same atoms of 1.59 Å. Figure reproduced from [24].

Very small deviations in these plots are usually only achieved after structural refinement by the use of PCSs [24, 55]. The deviations see in Figures 4.19-4.22 should therefore be expected given the RMSD of conformers of the MsrB1 structure.

Plots of PCSs for H^N vs. the hypothetical distance to the cobalt(II) ion were reported in section 4.8. As can be seen in Figures 4.23-4.25, the PCSs for the majority of residues follow a negative correlation with distance when conformers of the Zn-MsrB1 structure were used with zinc representing cobalt. This is consistent with the fact that the pseudocontact shift is proportional to distance in the negative power of three (PCS $\propto 1/r^3$), and immediately

shows that these residues constitute a part of Co-MsrB1 that has the same structure as in Zn-MsrB1. Ten residues are shown to diverge from this H^N-distance relationship in Figures 4.23 and 4.24, and these residues showed this divergence for all conformers (Figure 4.25). This could suggest that these residues had been erroneously assigned or that they represent a part of Co-MsrB1 for which the structure is different than in Zn-MsrB1. However, investigation of the location of these residues in the structure of Zn-MsrB1 revealed that the residues constitute two loops residing on opposite sides of the two structural CxxC-motifs demonstrated to be responsible for cobalt-coordination (Figure 5.5).
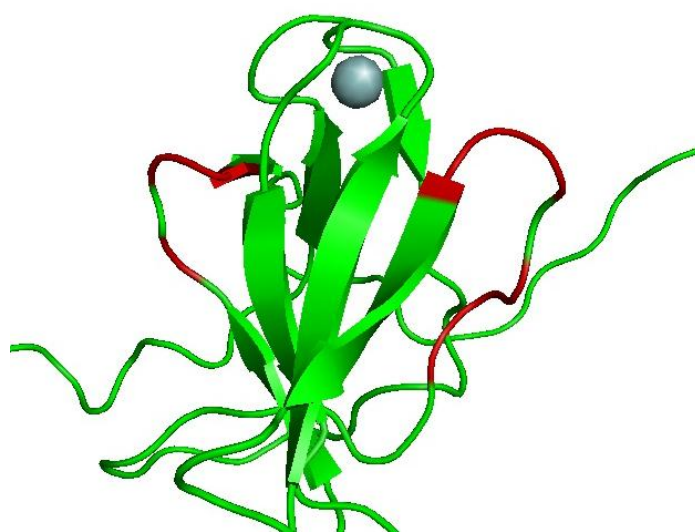
Figure 5.5: Two regions of Co-MsrB1 where PCS and H^N-metal binding site distance do not correlate. Residues for which H^N does not appear to follow a reverse proportional relationship with the distance to the demonstrated metal binding site of cobalt are marked red in the structure of Zn-MsrB1. Note that two residues interrupting the two stretches (marked green) have not been assigned.

Although a structural change cannot be completely ruled out, the explanation for the deviation from the distance-PCS-relationship of these residues is more likely to be local axial and rhombic components of magnetic susceptibility anisotropy [21], for which the magnetic field should have opposite signs and would lead to an increasingly positive PCS the further into these magnetic field components the H^N is positioned. Although the AnisoFit calculations show large deviations from the observed PCSs for these residues, the calculations are consistent with the observed near-zero or positive PCSs, and the deviations are likely due to the fact that these residues reside on unstructured loops with varying structure (Figure 5.5). PCSs-distance plots for the three conformers and the mean conformer together with mentioned observations thus provide strong evidence that the structures of Co- and Zn-MsrB1 are very similar, if not identical. In support of this, Figure 4.26 shows that a small change in the interatomic distances between the H^N of the assigned residues and the cobalt metal ion would have displayed significantly impaired PCS-distance proportionality. In the extreme scenario, as illustrated by Figure 4.27 where the metal

binding site is placed on the His-tag, the PCS-distance plot results in a "non-sense"-plot in which there is no proportionality between the distance to the cobalt metal ion and the pseudocontact shift. Figure 4.27 also further confirms that the His-tag cannot be responsible for uptake of cobalt as previously demonstrated (Dobrovolska et al., unpublished results).

As seen in Figures 4.19-4.22, the calculated PCSs with AnisoFit gave the best correlations with the observed PCSs for conformer 3. This indicates that conformer 3 is a better representation of the structure of Co-MsrB1 than the two conformers and the mean conformer that were evaluated. As can be seen in Figure 4.25, overlaying PCS-distance plots for all conformers draws a solid curve for the reversely distance-dependent residues, and converges to zero for the residues with a distance of approximately 35 Å or more from the cobalt metal ion. This curve, although somewhat broad, should represent the real distance-dependence of the PCSs of H^N. Comparing Figures 4.23, 4.24 and 4.25, the plot of conformer 3 appears to match this curve better than the plot of conformer 1. Particularly, the plot of conformer 1 converges to a PCS of +0.30 as distance increases, and these H^N represent residues of the flexible N-terminus. Although this could represent a calibration error of the NMR spectra, the PCSs for the C-terminus are zero (see Figure 4.24) showing that the positive PSCs of the N-terminus are not a calibrational mistake. In fact, these slightly positive PSCs indicate that the N-terminus spends significant time in proximity of the metal binding site, more specifically within the axial or rhombic magnetical susceptibility anisotropy components of the cobalt magnetic field. Indeed, this proximity is observed for conformer 3 (Figure 5.6) and is even more dramatic for some of the other calculated conformers which

Figure 5.6: The structure of conformer 3. The structure of conformer 3, which was shown to have better correlation between observed and calculated PCSs in AnisoFit than other conformers. The N-terminus (right, colored blue) is placed closer to the structural metal binding site than for the mean conformer.

were not used in AnisoFit calculations (specifically conformers 6, 10, 17 and 19). As catalytic activity depends on formation of a disulfide bridge between the catalytic Cys and the resolving Cys, the latter which resides on the N-terminus, it would not be surprising if the N-terminus tends towards being in proximity of the catalytic Cys, as it should increase the catalytic efficiency compared to if the N-terminus is completely unstructured (see also

Figure 1.34). In NOE-experiments, used to produce structural constraints in the published structure of MsrB1, detection of proximity is limited to protons within a distance of 4-5 Å[13]. If the N-terminus is infact positioned like that of conformer 3 or the similar conformers, it would not necessarily be detected by NOE. The PCSs contain within them structural constraints that should be used to investigate this further (see Further studies, section 5.10).

---

[13] Chemistry Research Laboratory, University of Oxford, NMR facility home page. Available online at http://www.chem.ox.ac.uk/spectroscopy/nmr/acropage.htm.

5.9 Biological significance of cobalt-substituted MsrB1

Since MsrB1 is a mammalian protein, the question naturally arises as to whether or not this substitution also happens in a mammalian cell. The metal-transport systems of *E. Coli* and other prokaryotes are well characterized, as described in section 1.2.7, but poorly understood for eukaryotes. Since higher eukaryotes are generally more complex on a cellular level than prokaryotes, metal transport and -insertion into polypeptides should be expected to equally complex, and thus would maybe not allow "mistakes" like cobalt-substitution to happen. However, evaluation of this question will only remain speculations until these systems are better characterized for mammals.

5.10 Further studies

As discussed in section 5.1, the *E. coli*-strain displayed exponential growth in the Zn-free M9-Co medium after successive Zn-free precultures, thus seemingly without access to free Zn(II). The Zn-proteome of *E. Coli* constitutes at least 3 % of the genome [56], including important enzymes such as DNA topoisomerase I, five tRNA synthases and RNA polymerase, and some of these proteins are present in high copy numbers [57, 58]. This raises the question of how the bacterium manages to live without Zn, and opens up for a very interesting possibility: might it be that Co and other metals take the place of Zn in these proteins? Given the seemingly promiscuous nature of protein Zn-binding sites *in vitro* with respect to a variety of other metals and evidence that substitution for other metals may produce functional enzymes [4], this does not seem unlikely. In fact, *in vivo* Co-substitution of RNA polymerase in *E. Coli* under zinc starvation has been reported [59]. The question of whether or not *E. Coli* generally can utilize other metals in zinc-binding sites when zinc is depleted is very interesting biologically, because zinc is generally regarded as a minimal requirement for growth[14]. In order to investigate this question, the genes of *E. Coli* known to be zinc-proteins should be cloned into an *E. Coli* expression vector (such as one used in this thesis) and overexpressed in Zn-free media supplemented with different trace metals (e.g. Co, Cu, Ni, etc). The purified proteins may then be investigated by UV-vis spectroscopy for uptake of metals. Structures of the proteins that do bind other metals could then also be studied in order to characterize which biochemical and/or structural properties allowing substitution.

It was argued in section 5.3 that metal-subsitution of MsrB1 *in vivo* is limited to cobalt due to the metal transport systems of *E. Coli*. To investigate this hypothesis it should be determined whether MsrB1 binds nickel and copper *in vitro*, and could be done by production of apo-MsrB1 (either by metal removal or restricting both cobalt and zinc in the growth medium) followed by dialysis against copper- and nickel, and the metal uptake would be checked through UV-vis spectroscopy. Although the attempts at reconstitution of Co-MsrB1 were unsuccessful in this study, important errors in the experimental conditions were pointed out in section 5.5 and successful reconstitution should thus be possible if these the experimental procedure is adjusted.

Possible pH-titrations were demonstrated in section 5.4. As was pointed out, the identity of the pH-titrated groups could not be established, and further experiments are needed to investigate this further. Mainly it should be determined whether titration curves are the result of loss of the cobalt metal ion (protonation of a cobalt-coordinating Cys) or titration of a residue in the vicinity of the cobalt-Cys complex. This should be done by determining the molar extinction coefficient (ε) for the same Co-MsrB1 solution first after dialysis in a

---

[14] Online textbook of bacteriology. Available online at http://textbookofbacteriology.net/nutgro.html (last visited 05.05.2012).

lower pH of one titration curve (e.g pH 5.5) and subsequently after dialysis in a higher pH of the same curve (pH 6.3). If ε is unchanged, this would show that there has been a loss of the cobalt metal ion, while reproduction of the titration points in Figure 5.2 would demonstrate the titrations represent groups outside the cobalt-Cys complex. If the experiment results in the latter, NMR-experiments should be performed on Co-MsrB1 to identify the titratable residues. These experiments would necessarily be large spectral width 1D ^{1}H experiments as the residues should be close to the cobalt-Cys complex and thus outside the spectral width of 2D- and 3D-experiments.

Conformer 3 was been demonstrated in section 5.8 to represent the structure of Co-MsrB1 better than the mean conformer, and calculations with AnisoFit should also be done for the remaining conformers (4-20). If conformers 6, 9, 17 and 19 also show better correlations than other conformers, this would represent strong evidence that the N-terminus is infact positioned towards the catalytical Cys as opposed to highly unstructured as shown in 1.34. As also discussed in section 5.8, the pseudocontact shift-data of Co-MsrB1 contain within them structural constraints. These data can be used to find the actual distance of the cobalt metal ion to the H^N-atoms and/or to obtain a general refinement of the already published structure of Zn-MsrB1 as described in [24].

6 Conclusion

In this study the cobalt(II)-substituted form of mammalian MsrB1, a native zinc protein, was biochemically and structurally characterized. This was done by expressing the Sec95Cys mutated His-tagged MsrB1 from *Mus musculus* in *E. coli* using a pET-21b plasmid construct in cobalt(II)dichloride ($CoCl_2$)-supplemented Lysegony broth (LB) and minimal (M9) medium followed by analysis of the purified protein by UV-visible absorption spectroscopy and high-resolution NMR spectroscopy.

The toxic effects of cobalt(II) on *E. coli* was demonstrated as decreasing growth rates with increasing concentrations of $CoCl_2$ in M9 medium. Uptake of cobalt(II) by MsrB1 was demonstrated for gene expression in all cobalt(II)-supplemented growth media by characteristic UV-vis absorption peaks for the purified protein. The absorption peaks were typical for that of a tetrahedrally four-coordinated cobalt-Cys complex, and it was demonstrated that these four Cys residues are the same residues that are responsible for zinc(II)-coordination in the native protein. Gene expression in nickel(II)-supplemented LB did not lead to Ni-MsrB1. The metal uptake characteristics of MsrB1 were argued to be the result of the intracellular metal concentrations in *E. Coli* created by the different metal transport systems.

The molar extinction coefficient (ε) was calculated from UV-vis absorption spectra of Co-MsrB1 in pH 4.9-11.5, and the resulting plot indicated two pH-titration curves at pH 4.9-6.3 and 7.2-10.9, the former with pK_a of about 5.6 and the latter with two pK_as of 8 and 10.1. It was argued that the titration curves either resulted from dissociation of the cobalt-Cys complex or titration of residues nearby the cobalt-Cys complex, and that further investigation is needed to identify the titratable groups.

The Co-MsrB1:Zn-MsrB1 ratios of MsrB1 expressed in M9 medium with 10 µM $CoCl_2$, LB with 50 µM $CoCl_2$ and LB with 10 µM $CoCl_2$ were determined by UV-vis spectroscopy to be about 0.2, 0.1 and 0.03, respectively. It was argued that the higher ratio from M9 medium is the result of a stress condition in which the metal efflux or influx systems suffer, leading to a higher intracellular concentration of cobalt(II).

Co- and Zn-MsrB1 was dialyzed against buffers containing chelating agents, and cobalt(II) was successfully removed at pH 5.0 and 5.5, while removal of zinc(II) was not achieved. This demonstrated that zinc(II) is more tightly bound by the metal-binding site than cobalt(II).

From 2D- and 3D-NMR experiments of Co-MsrB1, a 70% backbone-assignment and 50% side chain-assignment was achieved. The unassigned residues constituted two stretches of the polypeptide that included the two CxxC-motifs that coordinate cobalt(II), and several

strongly shifted signals were found from 350 to -80 ppm in large spectral width 1D ^{1}H NMR spectra which represented protons of the unassigned residues.

Pseudocontact shifts (PCSs) for H^N of Co-MsrB1 were determined using the published NMR-assignment of Zn-MsrB1. Using the observed PCSs a tensor was calculated by AnisoFit and PCSs were calculated and predicted for four structural conformers of the Zn-MsrB1 structure. The calculations gave the best correlations between observed and calculated PCSs for conformer 3, indicating that this conformer is a better representation of the structure of Co-MsrB1 than the other three conformers. Plots of PCSs vs. the hypothetical distance to the cobalt metal ion were made from the conformers of Zn-MsrB1, and demonstrated that the structure of Co- and Zn-MsrB1 are very similar. The N-terminus residues experience positive PCSs that indicate a proximity of these residues to the cobalt metal binding site, and it was argued that the PCSs should be further processed to investigate this possibility and to obtain a refinement of the published MsrB1 structure.

References

1.   Branden, C., Tooze, J., *Introduction to Protein Structure*. 2nd edition ed. 1999: Garland Publishing.

2.   Nelson, D.L., Cox, M. M., *Principles or Biochemistry.* Freeman, 2008. **5th ed.**

3.   Andreini, C., et al., *Metal ions in biological catalysis: from enzyme databases to general principles.* Journal of Biological Inorganic Chemistry, 2008. **13**(8): p. 1205-1218.

4.   Holm, R.H., P. Kennepohl, and E.I. Solomon, *ChemInform Abstract: Structural and Functional Aspects of Metal Sites in Biology.* ChemInform, 1997. **28**(10): p. no-no.

5.   Lippard, S.J., Berg, J. M., *Principles of Bioinorganic Chemistry.* 1994, Mill Valley, Calif.: University Science Books.

6.   Jaffe, E.K., et al., *Porphobilinogen synthase modification with methylmethanethiosulfonate. A protocol for the investigation of metalloproteins.* Journal of Biological Chemistry, 1984. **259**(8): p. 5032-5036.

7.   Klug, A. and J. Schwabe, *Protein motifs 5. Zinc fingers.* The FASEB Journal, 1995. **9**(8): p. 597-604.

8.   Hoffmann, K.M., et al., *Co(II)-substituted Haemophilus influenzae β-carbonic anhydrase: Spectral evidence for allosteric regulation by pH and bicarbonate ion.* Archives of Biochemistry and Biophysics, 2011. **511**(1–2): p. 80-87.

9.   Smith, A.T., et al., *Cobalt Cystathionine β-Synthase: A Cobalt-Substituted Heme Protein with a Unique Thiolate Ligation Motif.* Inorganic Chemistry, 2011. **50**(10): p. 4417-4427.

10.   Gessmann, R., et al., *The crystal structure of cobalt-substituted pseudoazurin from Alcaligenes faecalis.* Biopolymers, 2011. **95**(3): p. 202-207.

11.   Walsby, C.J., et al., *Cobalt-Substituted Zinc Finger 3 of Transcription Factor IIIA: Interactions with Cognate DNA Detected by 31P ENDOR Spectroscopy.* Journal of the American Chemical Society, 2003. **125**(25): p. 7502-7503.

12.   Kobayashi, M. and S. Shimizu, *Cobalt proteins.* European Journal of Biochemistry, 1999. **261**(1): p. 1-9.

13.   Arakawa, T., et al., *Structure of Thiocyanate Hydrolase: A New Nitrile Hydratase Family Protein with a Novel Five-coordinate Cobalt(III) Center.* Journal of Molecular Biology, 2007. **366**(5): p. 1497-1509.

14.   Waldron, K.J. and N.J. Robinson, *How do bacterial cells ensure that metalloproteins get the correct metal?* Nat Rev Micro, 2009. **7**(1): p. 25-35.

15.   Iwig, J.S., et al., *Ni(II) and Co(II) Sensing by Escherichia coli RcnR.* Journal of the American Chemical Society, 2008. **130**(24): p. 7592-7606.

16.   Kiemle, D., F.X. Webster, and R.M. Silverstein, *Spectrometric Identification of Organic Compounds.* 7th ed. ed. 2005: John Wiley & Sons.

17.   Bertini, I., C. Luchinat, and A. Rosato, *The solution structure of paramagnetic metalloproteins.* Progress in Biophysics and Molecular Biology, 1996. **66**(1): p. 43-80.

18.   Harrison, P.M. and P. Arosio, *The ferritins: molecular properties, iron storage function and cellular regulation.* Biochimica et biophysica acta, 1996. **1275**(3): p. 161-203.

19.   Abramowicz, D.A. and G.C. Dismukes, *Manganese proteins isolated from spinach thylakoid membranes and their role in O2 evolution: II. A binuclear manganese-containing 34 kilodalton protein, a probable component of the water dehydrogenase enzyme.* Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1984. **765**(3): p. 318-328.

20.   González-Iglesias, R., et al., *Prion Protein Interaction with Glycosaminoglycan Occurs with the Formation of Oligomeric Complexes Stabilized by Cu(II) Bridges.* Journal of Molecular Biology, 2002. **319**(2): p. 527-540.

21.   Koehler, J. and J. Meiler, *Expanding the utility of NMR restraints with paramagnetic compounds: Background and practical aspects.* 2011. **59**(4): p. 360-389.

22.   Bertini, I., et al., *The metal-binding properties of ovotransferrin. An investigation of cobalt(II) derivatives.* Journal of Biological Chemistry, 1986. **261**(3): p. 1139-1146.

23.   Banci, L., et al., *PSEUDYANA for NMR structure calculation of paramagnetic metalloproteins using torsion angle molecular dynamics.* Journal of Biomolecular NMR, 1998. **12**(4): p. 553-557.

24.   Banci, L., et al., *The use of pseudocontact shifts to refine solution structures of paramagnetic metalloproteins: Met80Ala cyano-cytochrome c as an example.* Journal of Biological Inorganic Chemistry, 1996. **1**(2): p. 117-126.

25.   Shi, Y., R.D. Beger, and J.M. Berg, *Metal binding properties of single amino acid deletion mutants of zinc finger peptides: studies using cobalt(II) as a spectroscopic probe.* Biophysical Journal, 1993. **64**(3): p. 749-753.

26.   Worthington, M.T., et al., *Metal Binding Properties and Secondary Structure of the Zinc-Binding Domain of Nup475.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(24): p. 13754-13759.

27.   Marcelo, G., *Oxyl radicals, redox-sensitive signalling cascades and antioxidants.* Cellular Signalling, 2007. **19**(9): p. 1807-1819.

28.   Turrens, J.F., *Mitochondrial formation of reactive oxygen species.* The Journal of Physiology, 2003. **552**(2): p. 335-344.

29.   Klaunig, J.E., et al., *Oxidative stress and oxidative damage in chemical carcinogenesis.* Toxicology and Applied Pharmacology, 2011. **254**(2): p. 86-99.

30. Buettner, G.R., *Superoxide Dismutase in Redox Biology: The roles of superoxide and hydrogen peroxide.* Anticancer Agents Med. Chem,, 2011. **11**(4): p. 341-346.

31. Berlett, B.S. and E.R. Stadtman, *Protein Oxidation in Aging, Disease, and Oxidative Stress.* Journal of Biological Chemistry, 1997. **272**(33): p. 20313-20316.

32. Lee, B.C. and V.N. Gladyshev, *The biological significance of methionine sulfoxide stereochemistry.* Free Radical Biology and Medicine, 2011. **50**(2): p. 221-227.

33. Lee, B.C., et al., *Functions and evolution of selenoprotein methionine sulfoxide reductases.* Biochimica et Biophysica Acta (BBA) - General Subjects, 2009. **1790**(11): p. 1471-1477.

34. Boschi-Muller, S., A. Gand, and G. Branlant, *The methionine sulfoxide reductases: Catalysis and substrate specificities.* Archives of Biochemistry and Biophysics, 2008. **474**(2): p. 266-273.

35. Zhang, X.-H. and H. Weissbach, *Origin and evolution of the protein-repairing enzymes methionine sulphoxide reductases.* Biological Reviews, 2008. **83**(3): p. 249-257.

36. Kauffmann, B., A. Aubry, and F. Favier, *The three-dimensional structures of peptide methionine sulfoxide reductases: current knowledge and open questions.* Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics, 2005. **1703**(2): p. 249-260.

37. Kim, H.-Y. and V.N. Gladyshev, *Different Catalytic Mechanisms in Mammalian Selenocysteine- and Cysteine-Containing Methionine-R-Sulfoxide Reductases.* PLoS Biol, 2005. **3**(12): p. e375.

38. Dobrovolska, O., et al., *Structural Insights into Interaction between Mammalian Methionine Sulfoxide Reductase B1 and Thioredoxin.* Journal of Biomedicine and Biotechnology, 2012. **2012**.

39. Elias S.J, A., *Selenoproteins—What unique properties can arise with selenocysteine in place of cysteine?* Experimental Cell Research, 2010. **316**(8): p. 1296-1303.

40. Allmang, C., L. Wurth, and A. Krol, *The selenium to selenoprotein pathway in eukaryotes: More molecular partners than anticipated.* Biochimica et Biophysica Acta (BBA) - General Subjects, 2009. **1790**(11): p. 1415-1423.

41. Kryukov, G.V., et al., *Characterization of Mammalian Selenoproteomes.* Science, 2003. **300**(5624): p. 1439-1443.

42. Aachmann, F.L., et al., *Insights into Function, Catalytic Mechanism, and Fold Evolution of Selenoprotein Methionine Sulfoxide Reductase B1 through Structural Analysis.* Journal of Biological Chemistry, 2010. **285**(43): p. 33315-33323.

43. Lee, T.-H. and H.-Y. Kim, *An anaerobic bacterial MsrB model reveals catalytic mechanisms, advantages, and disadvantages provided by selenocysteine and cysteine in reduction of methionine-R-sulfoxide.* Archives of Biochemistry and Biophysics, 2008. **478**(2): p. 175-180.

44.   Sal, L., et al., *NMR assignments of ^1H, ^{13}C and ^{15}N spectra of methionine sulfoxide reductase B1 from Mus musculus.* Biomolecular NMR Assignments, 2007. **1**(1): p. 131-133.

45.   Fischer, B., I. Sumner, and P. Goodenough, *Isolation, renaturation, and formation of disulfide bonds of eukaryotic proteins expressed in Escherichia coli as inclusion bodies.* Biotechnology and Bioengineering, 1993. **41**(1): p. 3-13.

46.   Tate, C.G., *Overexpression of mammalian integral membrane proteins for structural studies.* FEBS Letters, 2001. **504**(3): p. 94-98.

47.   May, S.W. and J.-Y. Kuo, *Preparation and properties of cobalt(II) rubredoxin.* Biochemistry, 1978. **17**(16): p. 3333-3338.

48.   Vasák, M. and J.H. Kägi, *Metal thiolate clusters in cobalt(II)-metallothionein.* Proceedings of the National Academy of Sciences, 1981. **78**(11): p. 6709-6713.

49.   Chen, X., M. Chu, and D.P. Giedroc, *Spectroscopic characterization of Co(II)-, Ni(II)-, and Cd(II)-substituted wild-type and non-native retroviral-type zinc finger peptides.* Journal of Biological Inorganic Chemistry, 2000. **5**(1): p. 93-101.

50.   Roth, I., et al., *2-Nitro-1,4-diaminobenzene-Functionalized Poly(vinyl amine)s as Water-Soluble UV-Vis-Sensitive pH Sensors.* Macromolecular Rapid Communications, 2006. **27**(3): p. 193-199.

51.   Bennett, B., *EPR of Cobalt-Substituted Zinc Enzymes*

in *Metals in Biology*, G. Hanson and L. Berliner, Editors. 2010, Springer New York. p. 345-370.

52.   Lim, B., et al., *RNase III Controls the Degradation of corA mRNA in Escherichia coli.* Journal of Bacteriology, 2012.

53.   Rensing, C., B. Mitra, and B.P. Rosen, *The zntA gene of Escherichia coli encodes a Zn(II)-translocating P-type ATPase.* Proceedings of the National Academy of Sciences, 1997. **94**(26): p. 14326-14331.

54.   Katta, V. and B.T. Chait, *Hydrogen/deuterium exchange electrospray ionization mass spectrometry: a method for probing protein conformational changes in solution.* Journal of the American Chemical Society, 1993. **115**(14): p. 6317-6321.

55.   Baig, I., et al., *Paramagnetism-Based Refinement Strategy for the Solution Structure of Human α-Parvalbumin†.* Biochemistry, 2004. **43**(18): p. 5562-5573.

56.   Katayama, A., et al., *Systematic search for zinc-binding proteins in Escherichia coli.* European Journal of Biochemistry, 2002. **269**(9): p. 2403-2413.

57.   Tse-Dinh, Y.C. and R.K. Beran-Steed, *Escherichia coli DNA topoisomerase I is a zinc metalloprotein with three repetitive zinc-binding domains.* Journal of Biological Chemistry, 1988. **263**(31): p. 15857-9.

58.   Outten, C.E., O'Halloran, and T. V., *Femtomolar Sensitivity of Metalloregulatory Proteins Controlling Zinc Homeostasis.* Science, 2001. **292**(5526): p. 2488-2492.

59. Speckhard, D.C., F.Y.H. Wu, and C.-W. Wu, *Role of the intrinsic metal in RNA polymerase from Escherichia coli. In vivo substitution of tightly bound zinc with cobalt.* Biochemistry, 1977. **16**(24): p. 5228-5234.

60. Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.* Analytical Biochemistry, 1976. **72**(1–2): p. 248-254.

Appendices

94

Appendix A: ProtParam of MsrB1 from ExPASy

The following parameters were obtained from ExPASy's ProtParam tool with the amino acid sequence of His-MsrB1U95C:

Sequence:

<pre>
        10   20   30   40   50   60
MSFCSFFGGE VFQNHFEPGV YVCAKCSYEL FSSHSKYAHS SPWPAFTETI HPDSVTKCPE

        70   80   90   100   110   120
KNRPEALKVS CGKCGNGLGH EFLNDGPKRG QSRFCIFSSS LKFVPKGKEA AASQGHLEHH

       124
HHHH
</pre>

Number of amino acids: 124

Molecular weight: 13806.5 g/mol

Theoretical pI: 8.19

Amino acid composition:

Ala (A)	7	5.6%
Arg (R)	3	2.4%
Asn (N)	4	3.2%
Asp (D)	2	1.6%
Cys (C)	7	5.6%
Gln (Q)	3	2.4%
Glu (E)	9	7.3%
Gly (G)	11	8.9%
His (H)	12	9.7%
Ile (I)	2	1.6%
Leu (L)	6	4.8%
Lys (K)	10	8.1%
Met (M)	1	0.8%
Phe (F)	11	8.9%
Pro (P)	8	6.5%
Ser (S)	15	12.1%
Thr (T)	3	2.4%
Trp (W)	1	0.8%
Tyr (Y)	3	2.4%
Val (V)	6	4.8%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 11
Total number of positively charged residues (Arg + Lys):   13

Atomic composition:

Carbon	C	615
Hydrogen	H	905
Nitrogen	N	175
Oxygen	O	175
Sulfur	S	8

Molecular formula: $C_{615}H_{905}N_{175}O_{175}S_8$
Total number of atoms: 1878

Extinction coefficients:
Extinction coefficients are in units of $M^{-1} cm^{-1}$, at 280 nm measured in water.
Extinction coefficient   10345
Abs 0.1% (=1 g/l)   0.749, assuming all pairs of Cys residues form cystines
Extinction coefficient   9970
Abs 0.1% (=1 g/l)   0.722, assuming all Cys residues are reduced

Estimated half-life:
The N-terminal of the sequence considered is M (Met).
The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
                              >20 hours (yeast, in vivo).
                              >10 hours (Escherichia coli, in vivo).

Instability index:
The instability index (II) is computed to be 40.30
This classifies the protein as unstable.

Aliphatic index: 44.84

Grand average of hydropathicity (GRAVY):   -0.566

Appendix B: Growth media, solutions and materials

B.1 Growth media, buffers and solutions

Unless specified, MilliQ water was used in preparation of media and solutions. All media and buffers were stored in 4°C unless when used immediately after preparation.

B.1.1 Lysogeny broth (LB)

Compound	Concentration (g/L)
Tryptone (manuf.?)	10
Yeast Extract (manuf?)	5
NaCl (sodium chloride)	10

The medium was autoclaved.

B.1.2 M9-medium (1/2 liter)

M9 salts (5x stock):

Compound	Concentration (g/L)
$Na_2HPO_4 \cdot 2\ H_2O$ (disodium hydrogen phosphate dihydrate)	30
KH_2PO_4 (potassium dihydrogen phosphate)	15
NaCl (sodium chloride)	2.5

The solution was split into 100 mL fractions, pH was adjusted to 7.0 with NaOH (6 M and 1 M) and autoclaved.

Nutrients, vitamins and metal salts:

Compound	Amount
D-Glucose	2 g
NH$_4$Cl (ammonium chloride)	0.5 g
"SL6" (10 mM Cu Mn ???)	5 mL
Ampicillin (100 mg/mL, 1000x stock)	0,5 mL
MgSO$_4$ (magnesium sulfate, 1 M)	1 mL
CaCl$_2$ (calcium chloride, 0.1 M)	0.5 mL
FeSO$_4$ (iron (II) sulfate, 0.01 M)	0.5 mL
Vitamine B1 (producer??, 0.1 %)	0.16 mL
CoCl$_2$ (cobalt (II) dichloride, 10 mM)	0.5 mL
ZnSO$_4$ (zinc (II) sulfate, 10 mM)	0.5 mL

The mixture was cold-sterilized using syringe mounted on a 0.22 µm filter, and added to 380 mL of autoclaved MilliQ water and 100 mL of M9 salt (5x stock) solution under sterile conditions.

CoCl$_2$ and ZnSO$_4$ are to be considered part of the M9 media, and were always prepared together with the other constituents listed. However, because their presence or absence are of key importance to the experiments, the CoCl$_2$ and ZnSO$_4$ concentrations are specified in the text together with the media.

B.1.3 Cell lysis, protein purification and NMR buffers

Unless specified, the pH of the buffers listed below was adjusted to 8.0 with NaOH (6M and 1M) and then cold-sterilized and degassed using a Millipore 0.22 m vacuum filter.

Cell lysis buffer:

Compound	Concentration
Na$_2$HPO$_4$ (disodium hydrogen phosphate)	50 mM
NaCl (sodium chloride)	400 mM
Polysorbate 20 (Tween-20, detergent)	0,01 % v/v
2-Mercaptoethanol	5 mM
Imidazole	*5 mM*
Complete protease inhibitor coctail	1 tablet per 50 mL

5 mM imidazole wash buffer (5mM buffer):

Compound	Concentration
Na_2HPO_4 (disodium hydrogen phosphate)	50 mM
NaCl (sodium chloride)	400 mM
Polysorbate 20 (Tween-20, detergent)	0,01 % v/v
2-Mercaptoethanol	5 mM
Imidazole	*5 mM*

10 mM imidazole wash buffer (10 mM buffer):

Compound	Concentration
Na_2HPO_4 (disodium hydrogen phosphate)	50 mM
NaCl (sodium chloride)	400 mM
Polysorbate 20 (Tween-20, detergent)	0,01 % v/v
2-Mercaptoethanol	5 mM
Imidazole	*10 mM*

20 mM imidazole wash buffer (20 mM buffer):

Compound	Concentration
Na_2HPO_4 (disodium hydrogen phosphate)	50 mM
NaCl (sodium chloride)	400 mM
Polysorbate 20 (Tween-20, detergent)	0,01 % v/v
2-Mercaptoethanol	5 mM
Imidazole	*20 mM*

250 mM imidazole elution buffer (250 mM buffer):

Compound	Concentration
Na_2HPO_4 (disodium hydrogen phosphate)	50 mM
NaCl (sodium chloride)	400 mM
Polysorbate 20 (Tween-20, detergent)	0,01 % v/v
2-Mercaptoethanol	5 mM
Imidazole	*250 mM*

NMR-buffer 1:

Compound	Concentration
Na_2HPO_4 (disodium hydrogen phosphate)	10 mM
NaCl (sodium chloride)	10 mM

The pH was adjusted to 5.5 and the solution was autoclaved.

NMR-buffer 2:

Compound	Concentration
Na_2HPO_4 (disodium hydrogen phosphate)	20 mM
NaCl (sodium chloride)	20 mM

The pH was adjusted to 5.5 and the solution was autoclaved.

Dialysis buffer 1:

Compound	Concentration
Na_2HPO_4 (disodium hydrogen phosphate)	50 mM
NaCl (sodium chloride)	400 mM
2-Mercaptoethanol	5 mM

The solution was prepared with 4 °C MilliQ water, and the pH was adjusted with HCL (1 M and 6 M) or NaOH (1 M and 6 M) to the pH stated in section 3. in the thesis. The solution was then cold-sterilized and degassed using a Millipore 0.22 μm vacuum filter.

Dialysis buffer 2:

Compound	Concentration
Na_2HPO_4 (disodium hydrogen phosphate)	50 mM
NaCl (sodium chloride)	30 mM
2-Mercaptoethanol	5 mM

The solution was prepared in the same way as described for dialysis buffer 1.

B.1.4 Bradford assay

Bradford assay is a water-based solution of red dye used for detection and quantification of protein. As the dye binds protein the color changes to blue [60]. In this thesis, Bradford assay was prepared with Bio-Rad Protein Assay diluted with MilliQ water (4:1) stored in a 50 mL plastic tube covered with aluminum foil. Before use the tube was gently shaken and aliquoted into eppendorf tubes (0.7 mL).

Appendix C: UV-vis absorption spectra for Co-MsrB1 in different pH

UV-vis absorption spectra for Co-MsrB1 in pH 4.9-11.5 used for determination of the molar extinction coefficient (ε) as described in section 3.2.4 are shown in Figure D.1.
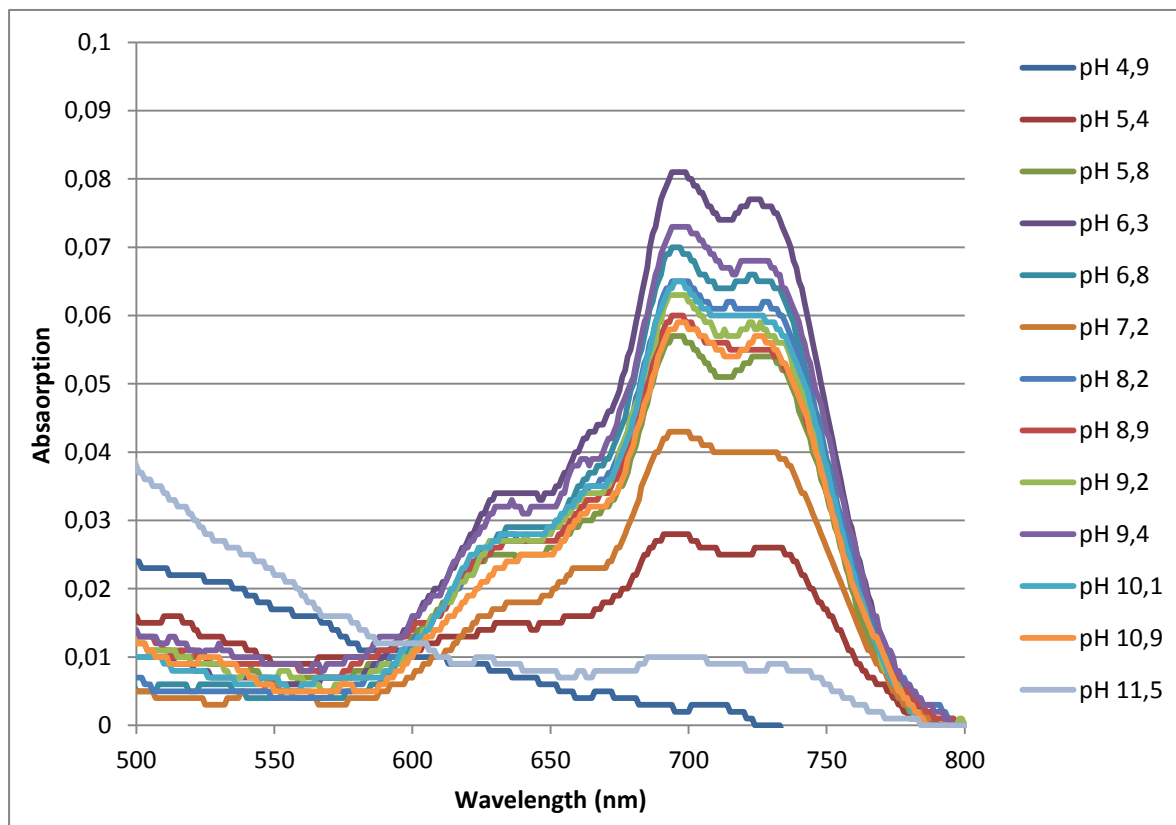
Figure D.1: UV-vis absorption spectra (500-800 nm) of Co-MsrB1 in different pH. UV-vis absorption spectra of Co-MsrB1in pH 4.9-11.5 in the 500-800 nm range overlaid. As discussed in section 5.4, no shift was observed in the wavelength of the absorption peaks, indicating that the number (or nature) of the ligands do not change with pH for Co-MsrB1. For pH 4.9, no characteristic absorption peaks were observed, showing that cobalt has been removed from the metal-binding site of all protein of the sample (apo-MsrB1).

Appendix D: Chemical shifts of Co-MsrB1

Chemical shifts of the backbone- and side chain-atoms from assignments are listed below.

Residue no.	Atom	Chemical shift (ppm)	Residue no.	Atom	Chemical shift (ppm)
4	CA	55.833	10	CB	26.371
4	CB	25.507	10	H	8.583
4	H	8.513	10	HA	4.403
4	HA	4.338	10	HB2	1.816
4	N	120.927	10	HB3	1.816
5	C	171.122	10	N	122.110
5	CA	55.801	11	CA	61.298
5	CB	60.912	11	CB	28.123
5	H	8.565	11	CG1	17.311
5	HA	4.372	11	CG2	15.947
5	HB2	3.804	11	H	8.073
5	HB3	3.762	11	HA	3.404
5	N	119.018	11	HB	1.430
6	C	172.574	11	HG1	0.235
6	CA	55.336	11	HG2	-0.032
6	CB	36.843	11	N	119.542
6	H	8.335	12	CA	52.840
6	HA	4.641	12	CB	34.543
6	HB2	3.052	12	H	7.543
6	HB3	2.980	12	HA	4.605
6	N	122.045	12	N	112.618
7	C	173.236	13	CA	59.245
7	CA	55.125	13	CB	23.726
7	CB	36.720	13	CG	29.389
7	H	8.548	13	H	7.896
7	HA	4.672	13	HA	2.751
7	HB2	3.137	13	HG2	1.896
7	HB3	3.053	13	HG3	1.896
7	N	122.532	13	N	125.419
8	C	171.682	14	CA	48.526
8	CA	42.931	14	CB	34.468
8	H	8.197	14	H	8.398
8	HA2	4.605	14	HA	4.606
8	HA3	4.605	14	HB2	2.602
8	N	110.741	14	HB3	2.300
9	C	172.562	14	N	114.860
9	CA	42.604	15	CA	53.675
9	H	8.258	15	CB	30.009
9	N	107.860	15	H	5.744
10	C	175.503	15	HA	3.798
10	CA	52.092	15	HB2	2.523

Residue no.	Atom	Chemical shift (ppm)	Residue no.	Atom	Chemical shift (ppm)
15	HB3	2.523	36	HA	4.162
15	N	119.424	36	HB2	0.712
16	C	171.305	36	HB3	0.712
16	CA	54.119	36	N	125.743
16	CB	36.071	37	CA	52.460
16	H	8.582	37	CB	37.669
16	HA	4.222	37	H	8.051
16	HB2	3.169	37	HA	4.679
16	HB3	3.169	37	HB2	2.501
16	N	126.185	37	HB3	2.501
17	CA	50.435	37	N	123.012
17	CB	27.181	38	CA	49.417
17	H	5.358	38	CB	14.855
17	HA	3.512	38	H	8.396
17	N	118.776	38	HA	4.197
19	CA	42.676	38	HB	0.808
19	H	8.393	38	N	124.069
19	HA2	4.007	39	CA	53.537
19	HA3	4.007	39	CB	31.200
19	N	110.016	39	H	7.880
20	CA	56.263	39	HA	4.202
20	CB	32.348	39	HB2	2.804
20	H	7.145	39	HB3	1.964
20	HA	2.905	39	N	124.756
20	HB	0.651	40	CA	56.022
20	HG1	-1.465	40	CB	59.673
20	N	119.070	40	H	6.737
21	CA	51.397	40	HA	3.763
21	CB	33.187	40	HB2	3.290
21	H	7.965	40	HB3	3.290
21	HA	2.332	40	N	117.760
21	HB2	0.244	41	CA	54.977
21	HB3	-0.501	41	CB	60.309
21	N	120.535	41	H	9.043
22	H	6.418	41	HA	4.196
35	CA	55.265	41	HB2	3.654
35	CB	61.921	41	HB3	3.654
35	H	7.631	41	N	121.652
35	HA	4.363	43	CA	54.297
35	HB2	3.805	43	CB	26.562
35	HB3	3.765	43	H	6.454
35	N	113.441	43	HA	4.423
36	CA	52.449	43	HB2	2.459
36	CB	32.175	43	HB3	2.459
36	H	7.246	43	N	112.992

Residue no.	Atom	Chemical shift (ppm)	Residue no.	Atom	Chemical shift (ppm)
45	CA	46.716	51	CA	51.242
45	CB	17.403	51	CB	29.853
45	H	7.363	51	H	9.139
45	HA	4.444	51	HA	6.606
45	HB	0.070	51	HB2	5.759
45	N	125.222	51	HB3	5.759
46	CA	51.219	51	N	116.500
46	CB	40.449	53	CA	52.210
46	H	7.414	53	CB	39.874
46	HA	5.086	53	H	10.727
46	HB2	2.204	53	HA	6.712
46	HB3	1.012	53	HB2	6.305
46	N	114.069	53	HB3	6.305
47	CA	59.945	53	N	116.000
47	CB	67.201	54	CA	66.671
47	CG2	21.062	54	CB	64.834
47	H	8.786	54	H	10.633
47	HA	4.648	54	HA	6.672
47	HB	4.071	54	N	119.522
47	HG2	0.794	55	CA	53.969
47	N	108.570	55	CB	29.296
48	CA	52.174	55	H	8.549
48	CB	28.927	55	HA	2.641
48	H	6.899	55	HB	0.158
48	HA	4.349	55	N	105.494
48	HB2	2.061	57	CA	50.323
48	HB3	1.749	57	CB	31.792
48	N	114.197	57	H	7.249
49	CA	57.173	57	HA	3.483
49	CB	67.438	57	N	119.350
49	CG2	18.789	58	CA	51.383
49	H	8.034	58	CB	25.255
49	HA	3.551	58	H	7.316
49	HB	4.265	58	HA	3.950
49	HG2	-0.591	58	HB2	1.449
49	N	110.231	58	HB3	1.449
50	CA	62.084	58	N	118.691
50	CB	36.307	61	C	173.948
50	CD1	12.276	61	CA	55.252
50	H	8.039	61	CB	28.359
50	HA	3.916	61	CG	21.342
50	HB	1.385	61	CD	29.389
50	HG1	1.093	61	CE	38.909
50	HD1	0.649	61	H	8.521
50	N	121.420	61	HA	3.505

Residue no.	Atom	Chemical shift (ppm)	Residue no.	Atom	Chemical shift (ppm)
61	HB2	1.288	67	N	119.046
61	HB3	1.288	82	CA	53.676
61	HG2	1.025	82	CB	35.302
61	HG3	1.000	82	H	8.650
61	HD2	1.684	82	HA	4.041
61	HD3	1.684	82	N	125.630
61	HE2	2.596	83	CA	52.558
61	HE3	2.596	83	CB	38.125
61	N	126.064	83	CD1	21.682
62	C	171.357	83	CD2	19.302
62	CA	50.945	83	H	7.946
62	CB	34.763	83	HA	3.623
62	H	8.697	83	HB2	1.147
62	HA	4.166	83	HB3	0.918
62	HB2	2.605	83	HG	2.570
62	HB3	2.499	83	HD1	0.327
62	N	115.939	83	HD2	0.112
63	C	169.161	83	N	125.612
63	CA	50.793	84	C	171.505
63	CB	27.989	84	CA	52.153
63	H	7.207	84	CB	34.651
63	HA	4.398	84	H	9.061
63	HB2	1.351	84	HA	3.917
63	HB3	1.032	84	HB2	2.683
63	N	118.151	84	HB3	2.439
65	C	171.195	84	N	115.885
65	CA	53.691	85	C	172.935
65	CB	25.017	85	CA	50.782
65	H	9.034	85	CB	42.142
65	HA	3.818	85	H	6.920
65	HB2	1.762	85	HA	4.818
65	HB3	1.762	85	HB2	2.030
65	N	116.066	85	HB3	1.773
66	C	171.561	85	N	116.210
66	CA	47.172	86	C	168.644
66	CB	18.480	86	CA	41.546
66	H	7.731	86	H	8.041
66	HA	4.695	86	HA2	4.814
66	HB	0.405	86	HA3	4.814
66	N	123.605	86	N	106.576
67	C	173.212	88	C	172.649
67	CA	49.500	88	CA	50.902
67	CB	39.440	88	CB	31.820
67	H	8.217	88	CG	21.857
67	HA	3.855	88	CD	26.404

Residue no.	Atom	Chemical shift (ppm)	Residue no.	Atom	Chemical shift (ppm)
88	CE	38.679	92	HB2	2.897
88	H	7.207	92	HB3	2.897
88	HA	4.188	92	N	119.888
88	HB2	1.418	93	C	172.857
88	HB3	1.141	93	CA	51.974
88	HG2	0.980	93	CB	31.681
88	HG3	0.980	93	H	7.451
88	HD2	1.317	93	HA	4.290
88	HD3	1.292	93	N	116.603
88	HE2	2.640	94	C	169.532
88	HE3	2.572	94	CA	54.834
88	N	115.263	94	CB	34.488
89	C	174.616	94	H	9.276
89	CA	55.208	94	HA	3.688
89	CB	26.224	94	HB2	1.814
89	CG	23.449	94	HB3	0.939
89	CD	40.384	94	N	125.639
89	H	8.263	95	C	169.538
89	HA	3.679	95	CA	54.119
89	HB2	1.443	95	CB	25.092
89	HB3	1.443	95	H	8.392
89	HG2	1.321	95	HA	3.898
89	HG3	1.321	95	N	126.799
89	HD2	2.937	96	C	170.432
89	HD3	2.937	96	CA	52.050
89	N	121.310	96	CB	35.281
90	C	171.970	96	CG1	21.062
90	CA	42.123	96	H	6.740
90	H	8.875	96	HA	3.460
90	HA2	3.933	96	HB	0.127
90	HA3	3.510	96	HG1	-0.303
90	N	114.322	96	HG1	-1.733
91	C	173.937	96	HG2	-1.207
91	CA	53.048	96	N	128.960
91	CB	28.197	97	C	176.010
91	H	8.031	97	CA	54.775
91	HA	4.202	97	CB	35.738
91	HB2	1.772	97	H	7.623
91	HB3	1.772	97	HA	4.774
91	N	120.064	97	HB2	3.573
92	C	168.609	97	HB3	3.536
92	CA	54.548	97	N	121.144
92	CB	62.049	98	CA	62.272
92	H	8.996	98	CB	60.684
92	HA	4.272	98	H	11.846

Residue no.	Atom	Chemical shift (ppm)	Residue no.	Atom	Chemical shift (ppm)
98	N	126.436	104	HG1	-0.595
99	C	175.582	104	HG2	-1.725
99	CA	58.939	104	N	129.551
99	CB	61.435	106	C	173.641
99	H	10.289	106	CA	53.702
99	HA	5.497	106	CB	30.235
99	HB2	6.122	106	CG	21.517
99	HB3	6.122	106	CD	26.176
99	N	114.182	106	CE	39.020
100	CA	60.947	106	H	7.475
100	CB	63.940	106	HA	3.618
100	H	9.496	106	HB2	1.171
100	HA	-3.185	106	HB3	0.972
100	N	120.285	106	HG2	0.936
101	C	171.122	106	HG3	0.936
101	CA	48.431	106	HD2	1.249
101	CB	39.929	106	HD3	1.249
101	H	8.235	106	HE2	2.565
101	HA	4.183	106	HE3	2.565
101	HB2	1.902	106	N	121.966
101	HB3	1.427	107	C	170.962
101	N	120.834	107	CA	41.963
102	C	168.855	107	H	8.505
102	CA	49.683	107	HA2	3.480
102	CB	30.128	107	HA3	3.480
102	H	6.206	107	N	112.483
102	HA	2.652	108	C	173.686
102	HB2	-1.972	108	CA	53.166
102	HB3	-2.504	108	CB	29.584
102	N	128.768	108	CG	21.569
103	C	169.440	108	CD	29.105
103	CA	52.089	108	CE	39.022
103	CB	37.022	108	H	8.024
103	H	7.363	108	HA	3.874
103	HA	2.588	108	HB2	1.235
103	HB2	1.351	108	HB3	1.207
103	HB3	0.776	108	HG2	0.939
103	N	121.711	108	HG3	0.939
104	C	168.964	108	HD2	1.351
104	CA	54.612	108	HD3	1.351
104	CB	27.989	108	HE2	2.536
104	CG1	16.402	108	HE3	2.536
104	H	6.934	108	N	120.346
104	HA	2.642	109	C	173.394
104	HB	0.173	109	CA	53.664

Residue no.	Atom	Chemical shift (ppm)	Residue no.	Atom	Chemical shift (ppm)
109	CB	26.952	114	HB3	1.660
109	CG	33.217	114	HG2	2.008
109	H	8.488	114	HG3	1.949
109	HA	3.753	114	N	121.621
109	HB2	1.655	115	C	171.076
109	HB3	1.562	115	CA	42.738
109	HG2	1.909	115	H	8.414
109	HG3	1.909	115	HA2	3.692
109	N	121.711	115	HA3	3.692
110	C	174.717	115	N	109.315
110	CA	49.650	116	C	172.006
110	CB	16.196	116	CA	52.933
110	H	8.259	116	CB	26.556
110	HA	3.829	116	H	8.312
110	HB	0.950	116	HA	4.398
110	N	125.163	116	HB2	2.910
111	C	174.876	116	HB3	2.910
111	CA	49.721	116	N	118.453
111	CB	16.181	117	C	174.216
111	H	8.119	117	CA	52.362
111	HA	3.900	117	CB	39.406
111	HB	1.022	117	CD1	21.403
111	N	122.808	117	CD2	52.362
112	C	175.003	117	H	8.247
112	CA	49.801	117	HA	4.083
112	CB	15.797	117	HB2	1.196
112	H	8.147	117	HB3	1.196
112	HA	3.957	117	HD1	0.659
112	HB	1.079	117	HD2	0.601
112	N	122.760	117	N	123.276
113	C	171.891	118	C	173.286
113	CA	55.669	118	CA	53.714
113	CB	60.804	118	CB	27.405
113	H	8.159	118	CG	33.244
113	HA	4.111	118	H	8.417
113	HB2	3.613	118	HA	3.973
113	HB3	3.566	118	HB2	1.709
113	N	114.573	118	HB3	1.662
114	C	173.651	118	HG2	2.011
114	CA	53.172	118	HG3	1.959
114	CB	26.499	118	N	121.451
114	CG	33.242	119	C	171.744
114	H	8.331	119	CA	52.935
114	HA	3.992	119	CB	26.919
114	HB2	1.856	119	H	8.430

Residue no.	Atom	Chemical shift (ppm)
119	HA	4.430
119	HB2	2.831
119	HB3	2.831
119	N	119.821

Appendix E: Calculated and observed PCSs

Below, calculated pseudocontact shifts (C. Pcs) and observed pseudocontact shift (O. Pcs) of [1]H-signals for backbone H-atoms of Co-MsrB1 are listed together with residue number to which they belong. The conformers on which the calculations are based are specified for each dataset. The calculations were done by Rebecca del Conte, CERM, Italy, and are based on the third conformer of the published structure for Zn-MsrB1 (PDB ID: 2KV1).

Conformer 1:

C. pcs	O. pcs	Res.nr/atom	C. pcs	O. pcs	Res.nr/atom
−0.0028430198	0.31	4 H	−0.160408777	−0.17	62 H
−0.00985499018	0.31	5 H	−0.161073289	−0.22	63 H
−0.014708244	0.33	6 H	−0.164363157	−0.18	65 H
−0.018880153	0.32	7 H	−0.210389794	−0.31	66 H
−0.0290837693	0.35	8 H	−0.350240195	−0.49	67 H
−0.0236261324	0.33	9 H	−0.28685305	−0.50	82 H
−0.0309004425	0.18	10 H	−0.271518452	−0.33	83 H
−0.0432493136	0.13	11 H	−0.158671455	−0.20	84 H
−0.066744779	−0.19	12 H	−0.142530215	−0.22	85 H
−0.0858134029	−0.18	13 H	−0.116666344	−0.19	86 H
−0.118751803	−0.18	14 H	−0.0594748074	−0.10	88 H
−0.156546544	−0.32	15 H	−0.0633882237	−0.06	89 H
−0.207936163	−0.30	16 H	−0.073278607	−0.07	90 H
−0.24978709	−0.42	17 H	−0.0892602245	−0.09	91 H
−0.420954514	−0.43	19 H	−0.14011099	−0.21	92 H
−0.886350788	−0.92	20 H	−0.210307029	−0.37	93 H
−1.38408479	−1.60	21 H	−0.31586829	−0.58	94 H
0.00674015941	−0.33	35 H	−0.480083486	−0.80	95 H
−0.137921484	−0.38	36 H	−0.476383703	−0.77	96 H
−0.194887999	−0.39	37 H	−0.121841192	−0.42	97 H
−0.137277041	−0.19	38 H	1.14199751	0.93	99 H
−0.181034168	−0.20	39 H	1.40890556	1.35	100 H
−0.121201947	−0.12	40 H	−0.238722278	0.13	101 H
−0.138081281	−0.10	41 H	−1.31131647	−1.65	103 H
−0.0936473001	−0.13	43 H	−1.44210829	−1.62	104 H
−0.267382361	−0.49	45 H	−0.446246972	−0.42	106 H
−0.374807178	−0.67	46 H	−0.309387334	−0.24	107 H
−0.204044851	−0.40	47 H	−0.227920146	−0.22	108 H
−0.14496249	−0.42	48 H	−0.191537623	−0.16	109 H
0.0162494926	−0.12	49 H	−0.15366485	−0.14	110 H
0.428652592	0.32	50 H	−0.147504248	−0.12	111 H
0.65365855	1.74	51 H	−0.135484588	−0.07	112 H
3.33101866	2.66	53 H	−0.0549488127	−0.04	113 H
5.48302436	2.91	54 H	−0.00579862785	0.01	114 H
3.41204105	1.19	55 H	0.0390838806	0.04	115 H
−1.31754285	−1.36	57 H	0.0441675471	0.02	116 H
−1.09071195	−1.29	58 H	0.0790922873	0.03	117 H
−0.224888879	−0.26	61 H	0.092957688	0.07	118 H

Conformer 2:

C. pcs	O. pcs	Res.nr/atom	C. pcs	O. pcs	Res.nr/atom
0.00342634721	0.31	4 H	−0.130442642	−0.22	63 H
0.000714168721	0.31	5 H	−0.164550273	−0.18	65 H
−0.0015779646	0.33	6 H	−0.216542498	−0.31	66 H
−0.0142952172	0.32	7 H	−0.384681952	−0.49	67 H
−0.0191020837	0.35	8 H	−0.409182996	−0.50	82 H
−0.0223958919	0.33	9 H	−0.290561312	−0.33	83 H
−0.0380867788	0.18	10 H	−0.162685855	−0.20	84 H
−0.0750231699	0.13	11 H	−0.150744632	−0.22	85 H
−0.101265333	−0.19	12 H	−0.117180213	−0.19	86 H
−0.100420394	−0.18	13 H	−0.0437859063	−0.10	88 H
−0.112404304	−0.18	14 H	−0.0417852609	−0.06	89 H
−0.154199439	−0.32	15 H	−0.0331911645	−0.07	90 H
−0.164518059	−0.30	16 H	−0.0615640766	−0.09	91 H
−0.212428685	−0.42	17 H	−0.168199399	−0.21	92 H
−0.426740198	−0.43	19 H	−0.302684593	−0.37	93 H
−0.783185497	−0.92	20 H	−0.475213444	−0.58	94 H
−1.29884947	−1.60	21 H	−0.802646922	−0.80	95 H
−0.138196955	−0.33	35 H	−1.01700662	−0.77	96 H
−0.250543476	−0.38	36 H	−0.325409183	−0.42	97 H
−0.298613087	−0.39	37 H	1.12217245	0.93	99 H
−0.178792144	−0.19	38 H	1.95748418	1.35	100 H
−0.228150147	−0.20	39 H	−0.26497848	0.13	101 H
−0.157447103	−0.12	40 H	−1.16613481	−1.65	103 H
−0.201301286	−0.10	41 H	−1.24488586	−1.62	104 H
−0.291154741	−0.13	43 H	−0.394219886	−0.42	106 H
−0.344670941	−0.49	45 H	−0.25876365	−0.24	107 H
−0.562693037	−0.67	46 H	−0.199194214	−0.22	108 H
−0.368191559	−0.40	47 H	−0.157953649	−0.16	109 H
−0.314290116	−0.42	48 H	−0.135002248	−0.14	110 H
−0.0648887508	−0.12	49 H	−0.10468698	−0.12	111 H
0.257921232	0.32	50 H	−0.0806674887	−0.07	112 H
2.02638575	1.74	51 H	−0.0597803962	−0.04	113 H
2.75190999	2.66	53 H	−0.0499777758	0.01	114 H
5.85138256	2.91	54 H	−0.0406865843	0.04	115 H
3.80751865	1.19	55 H	−0.0314520396	0.02	116 H
−1.01031457	−1.36	57 H	−0.0262622087	0.03	117 H
−0.936389168	−1.29	58 H	−0.0251275433	0.07	118 H
−0.16837511	−0.26	61 H	−0.0234568214	0.02	119 H
−0.113809885	−0.17	62 H			

Conformer 3:

C. pcs	O. pcs	Res.nr/atom	C. pcs	O. pcs	Res.nr/atom
0.199997677	0.31	4 H	−0.209568905	−0.22	63 H
0.168850116	0.31	5 H	−0.224254819	−0.18	65 H
0.200242636	0.33	6 H	−0.283684539	−0.31	66 H
0.169302578	0.32	7 H	−0.489096981	−0.49	67 H
0.136069534	0.35	8 H	−0.500864442	−0.50	82 H
0.0834487923	0.33	9 H	−0.296881892	−0.33	83 H
0.0407263208	0.18	10 H	−0.190709615	−0.20	84 H
−0.0377483386	0.13	11 H	−0.17431826	−0.22	85 H
−0.0718318499	−0.19	12 H	−0.143519423	−0.19	86 H
−0.106139279	−0.18	13 H	−0.078483555	−0.10	88 H
−0.141230775	−0.18	14 H	−0.0836379784	−0.06	89 H
−0.211243348	−0.32	15 H	−0.0730925994	−0.07	90 H
−0.212528501	−0.30	16 H	−0.113031069	−0.09	91 H
−0.248036462	−0.42	17 H	−0.231985159	−0.21	92 H
−0.493903862	−0.43	19 H	−0.345812948	−0.37	93 H
−0.9352838	−0.92	20 H	−0.557146627	−0.58	94 H
−1.77715058	−1.60	21 H	−0.910393805	−0.80	95 H
−0.237186051	−0.33	35 H	−0.811665376	−0.77	96 H
−0.370181474	−0.38	36 H	−0.584378818	−0.42	97 H
−0.417094279	−0.39	37 H	1.23237968	0.93	99 H
−0.265711124	−0.19	38 H	1.30925381	1.35	100 H
−0.21456491	−0.20	39 H	0.0634579886	0.13	101 H
−0.139597711	−0.12	40 H	−1.06091352	−1.65	103 H
−0.150782548	−0.10	41 H	−1.41436296	−1.62	104 H
−0.169172776	−0.13	43 H	−0.455323617	−0.42	106 H
−0.383449645	−0.49	45 H	−0.302448666	−0.24	107 H
−0.67875337	−0.67	46 H	−0.226231144	−0.22	108 H
−0.45599648	−0.40	47 H	−0.170641008	−0.16	109 H
−0.374509632	−0.42	48 H	−0.157032683	−0.14	110 H
−0.0770550308	−0.12	49 H	−0.118024517	−0.12	111 H
0.425501527	0.32	50 H	−0.111381232	−0.07	112 H
1.66513559	1.74	51 H	−0.0922770195	−0.04	113 H
1.9091417	2.66	53 H	−0.0647290544	0.01	114 H
3.31971214	2.91	54 H	−0.0528967525	0.04	115 H
2.64096493	1.19	55 H	−0.0394055341	0.02	116 H
−1.21760521	−1.36	57 H	−0.0297679023	0.03	117 H
−1.27574734	−1.29	58 H	−0.0217724308	0.07	118 H
−0.279307441	−0.26	61 H	−0.00932395868	0.02	119 H
−0.192590269	−0.17	62 H			

Mean (average) conformer:

C. pcs	O. pcs	Res.nr/atom	C. pcs	O. pcs	Res.nr/atom
-0.0970468708	0.31	4 H	-0.133212668	-0.22	63 H
-0.110593082	0.31	5 H	-0.161377394	-0.18	65 H
-0.119031547	0.33	6 H	-0.209084871	-0.31	66 H
-0.123938848	0.32	7 H	-0.380813527	-0.49	67 H
-0.145232333	0.35	8 H	0.406056689	-0.50	82 H
-0.141516791	0.33	9 H	-0.286220072	-0.33	83 H
-0.148211078	0.18	10 H	-0.2016252	-0.20	84 H
-0.141289759	0.13	11 H	-0.173776017	-0.22	85 H
-0.149870523	-0.19	12 H	-0.15044034	-0.19	86 H
-0.145343491	-0.18	13 H	-0.0957026689	-0.10	88 H
-0.136754858	-0.18	14 H	-0.0940917014	-0.06	89 H
-0.150131477	-0.32	15 H	-0.0995799768	-0.07	90 H
-0.147034279	-0.30	16 H	-0.123805063	-0.09	91 H
-0.18830249	-0.42	17 H	-0.192929613	-0.21	92 H
-0.331184258	-0.43	19 H	-0.29647379	-0.37	93 H
-0.652623036	-0.92	20 H	-0.454934792	-0.58	94 H
-1.11705459	-1.60	21 H	-0.693908479	-0.80	95 H
-0.040481515	-0.33	35 H	-0.686225411	-0.77	96 H
-0.189997845	-0.38	36 H	-0.297456641	-0.42	97 H
-0.247973116	-0.39	37 H	1.24813007	0.93	99 H
-0.1494601	-0.19	38 H	1.84209324	1.35	100 H
-0.188565903	-0.20	39 H	-0.281741507	0.13	101 H
-0.133549264	-0.12	40 H	-1.40969507	-1.65	103 H
-0.159458088	-0.10	41 H	-1.11982927	-1.62	104 H
-0.150868058	-0.13	43 H	-0.318315537	-0.42	106 H
-0.301090262	-0.49	45 H	-0.200656235	-0.24	107 H
-0.521025207	-0.67	46 H	-0.165089476	-0.22	108 H
-0.326791189	-0.40	47 H	-0.133491655	-0.16	109 H
-0.270218033	-0.42	48 H	-0.122194327	-0.14	110 H
-0.123182027	-0.12	49 H	-0.102281716	-0.12	111 H
0.53772337	0.32	50 H	-0.0885590217	-0.07	112 H
1.58704153	1.74	51 H	-0.0724751741	-0.04	113 H
2.51256246	2.66	53 H	-0.061403288	0.01	114 H
3.22856579	2.91	54 H	-0.046752295	0.04	115 H
-0.458503953	1.19	55 H	-0.039768532	0.02	116 H
-1.31708947	-1.36	57 H	-0.0307949345	0.03	117 H
-0.907614486	-1.29	58 H	-0.0274572016	0.07	118 H
-0.167891592	-0.26	61 H	-0.0229719416	0.02	119 H
-0.118690976	-0.17	62 H			

Appendix F: Calculated PCSs for Co-MsrB1 and H^N-Zn distances for conformer 1

Table F.1: Chemical shifts of H^N, calculated pseudocontact shifts and H^N-Zn distance for conformer 1. The chemical shifts for H^N of Co-MsrB1 and Zn-MsrB1 are listed together with the calculated pseudocontact shift for Co-MsrB1 and the calculated H^N-Zn distances for the respective residues in conformer 1.

δHn Co-MsrB1, ppm	Res.nr	Res.type	δHn Zn-MsrB1, ppm	ΔδHn = PCS, ppm	Hn-Zn distance, Å
8,05	37	TYR	8,44	-0,39	21,4
8,4	38	ALA	8,59	-0,19	24,9
7,88	39	HIS	8,08	-0,2	24,7
6,74	40	SER	6,86	-0,12	28,2
9,04	41	SER	9,14	-0,1	26,0
	42	PRO			
6,46	43	TRP	6,59	-0,13	24,5
	44	PRO			
7,36	45	ALA	7,85	-0,49	20,4
7,41	46	PHE	8,08	-0,67	17,7
8,79	47	THR	9,19	-0,4	19,7
6,9	48	GLU	7,32	-0,42	18,8
8,03	49	THR	8,15	-0,12	16,7
8,04	50	ILE	7,72	0,32	12,8
9,14	51	HIS	7,40	1,74	13,2
	52	PRO			
10,73	53	ASP	8,07	2,66	11,4
10,63	54	SER	7,72	2,91	9,3
8,55	55	VAL	7,36	1,19	8,4
	56	THR			
7,249	57	LYS	8,61	-1,361	12,9
7,32	58	CYS	8,61	-1,29	14,6
	59	PRO			
	60	GLU			20,6
8,52	61	LYS	8,78	-0,26	22,8
8,7	62	ASN	8,87	-0,17	24,5
7,21	63	ARG	7,43	-0,22	24,7
	64	PRO			
9,03	65	GLU	9,21	-0,18	25,2
7,73	66	ALA	8,04	-0,31	23,4
8,22	67	LEU	8,71	-0,49	19,5
	68	LYS			
	69	VAL			
	70	SER			
	71	CYS			
	72	GLY			
	73	LYS			
	74	CYS			
	75	GLY			
	76	ASN			

δHn Co-MsrB1, ppm	Res.nr	Res.type	δHn Zn-MsrB1, ppm	ΔδHn = PCS, ppm	Hn-Zn distance, Å
	77	GLY			
	78	LEU			
	79	GLY			
	80	HIS			
	81	GLU			
8,65	82	PHE	9,15	-0,5	19,8
7,95	83	LEU	8,28	-0,33	20,9
9,06	84	ASN	9,26	-0,2	23,9
6,92	85	ASP	7,14	-0,22	25,5
8,04	86	GLY	8,23	-0,19	26,5
	87	PRO			
7,21	88	LYS	7,31	-0,1	30,6
8,26	89	ARG	8,32	-0,06	30,3
8,88	90	GLY	8,95	-0,07	29,1
8,03	91	GLN	8,12	-0,09	27,8
9	92	SER	9,21	-0,21	24,7
7,45	93	ARG	7,82	-0,37	21,5
9,28	94	PHE	9,86	-0,58	18,2
8,39	95	CYS	9,19	-0,8	16,2
6,74	96	ILE	7,51	-0,77	15,0
7,62	97	PHE	8,04	-0,42	12,0
11,85	98	SER			14,1
10,29	99	SER	9,36	0,93	12,6
9,5	100	SER	8,15	1,35	10,1
8,23	101	LEU	8,10	0,13	9,0
	102	LYS			
7,36	103	PHE	9,01	-1,65	13,1
6,93	104	VAL	8,55	-1,62	14,1
	105	PRO			
7,48	106	LYS	7,90	-0,42	20,6
8,5	107	GLY	8,74	-0,24	23,4
8,02	108	LYS	8,24	-0,22	26,2
8,49	109	GLU	8,65	-0,16	27,2
8,26	110	ALA	8,40	-0,14	28,1
8,12	111	ALA	8,24	-0,12	25,8
8,15	112	ALA	8,22	-0,07	23,4
8,16	113	SER	8,20	-0,04	25,6
8,33	114	GLN	8,32	0,01	29,4
8,41	115	GLY	8,37	0,04	28,0
8,31	116	HIS	8,29	0,02	31,0
8,25	117	LEU	8,22	0,03	29,3
8,42	118	GLU	8,35	0,07	31,5
8,43	119	HIS	8,41	0,02	31,0